Evaluation et Méthode Scientifique

Martin Vielvoye 2021

Performance, précision et évaluation

Comment évaluer les performances d'un modèle ?

Régression Linéaire

- Pour la régression linéaire, la moyenne d'erreur sur le jeu de Test permet de donner une valeur cohérente par rapport à la valeur prédite.

$$J_{test}(\theta) = \frac{1}{2m_{test}} \sum_{i=1}^{m_{test}} (\hat{y}_{test}^{(i)} - y_{test}^{(i)})^2$$

- Exemple :
  - Si un prédicteur de prix de maison à un coût de `1,000,000` sur la donnée de test, alors on sait qu'en moyenne, le modèle à une erreur de `±1000 $` sur le vrai prix de la maison.

Comment évaluer les performances d'un modèle ?

Classification

- Pour la classification, on peut évaluer le nombre de mauvaises classifications que le modèle a fait et on calcule le pourcentage d'erreur.

$$E_{test} = \frac{1}{m_{test}} \sum_{i=1}^{m_{test}} err(\widehat{y}^{(i)}, y^{(i)})$$

- Exemple : Si un modèle classifie correctement `55` photos sur `70` photos, alors on peut calculer: $\frac{55}{70} = 0{,}785$ et dire que notre modèle a une performance de `78,5%`.

Analyse de Résultat

Matrice de Confusion

On peut diviser les résultats de classifications et analyser à quels endroits le modèle réussi et échoue.

2 groupes de catégories :

- <u>Vrai Positif</u> et <u>Vrai Négatif</u>
  - Les valeurs correctement classifiées
- <u>Faux Positif</u> et <u>Faux Négatif</u>
  - Les valeurs mal classifiées

Objectif / Prédiction	1	0
$\hat{1}$	Vrai Positif	Faux Positif
$\hat{0}$	Faux Négatif	Vrai Négatif

Analyse de Résultat

Matrice de Confusion

1000	h(X) = 0
995 -> 0 5 -> 1	995/1000 = 99.5%

Il est important de vérifier que le jeu de données possède une quantité raisonnable de chaque donnée (0 et 1).

Exemple : Un jeu de donnée, tentant de prédire si un patient a une maladie rare, ne possède que 0,05% de cas positif. On écrit un algorithme qui prédît un $\hat{0}$ dans tout les cas et on observe une performance de 99,95%.

➡ Il nous faut donc analyser plus précisément la matrice de confusion.

Actuelle / Prédiction	1 5	0 995
$\hat{1}$ 0	Vrai Positif - **0**%	Faux Positif - **0**%
$\hat{0}$ 1000	Faux Négatif - **0**%	Vrai Négatif - **100**%

Analyse de Résultat

Matrice de Confusion

	1	0
$\hat{1}$	Vrai Positif	Faux Positif
$\hat{0}$	Faux Négatif	Vrai Négatif

- **Ex 1** : Quelles sont les groupes les plus important dans la prédiction des maladies graves qui inclut une médication légère et sans danger?
  - Faux Négatif et Vrai Positif
- **Ex 2** : Quelles sont les groupes les plus importants dans la prédiction de maladies non graves mais qui inclut des médications lourdes et dangereuse?
  - Faux Positif et Vrai Positif
- **Ex 3** : Quelles sont les groupes les plus importants dans la prédiction de maladies graves et qui inclut des médications lourdes et dangereuse?
  - Faux Positif, Faux Négatif et Vrai Positif

Analyse de Résultat

Précision et Recall

	1	0
$\hat{1}$	Vrai Positif	Faux Positif
$\hat{0}$	Faux Négatif	Vrai Négatif

- Deux modèles qui ont la même score de 97% mais qui n'ont pas les mêmes ratios de Faux Négatifs et de Faux Positifs peuvent avoir des comportement très différents.

$$Precision = \frac{VP}{VP + FP}$$

$$Recall = \frac{VP}{VP + FN}$$

Analyse de Résultat

F-Score

	1	0
$\hat{1}$	Vrai Positif	Faux Positif
$\hat{0}$	Faux Négatif	Vrai Négatif

- On peut utiliser le score de Précision et de Recall pour créer un score qui n'est haut **que** quand la Précision **ET** le Recall sont élevés.

$$Precision = \frac{VP}{VP + FP}$$

$$Recall = \frac{VP}{VP + FN}$$

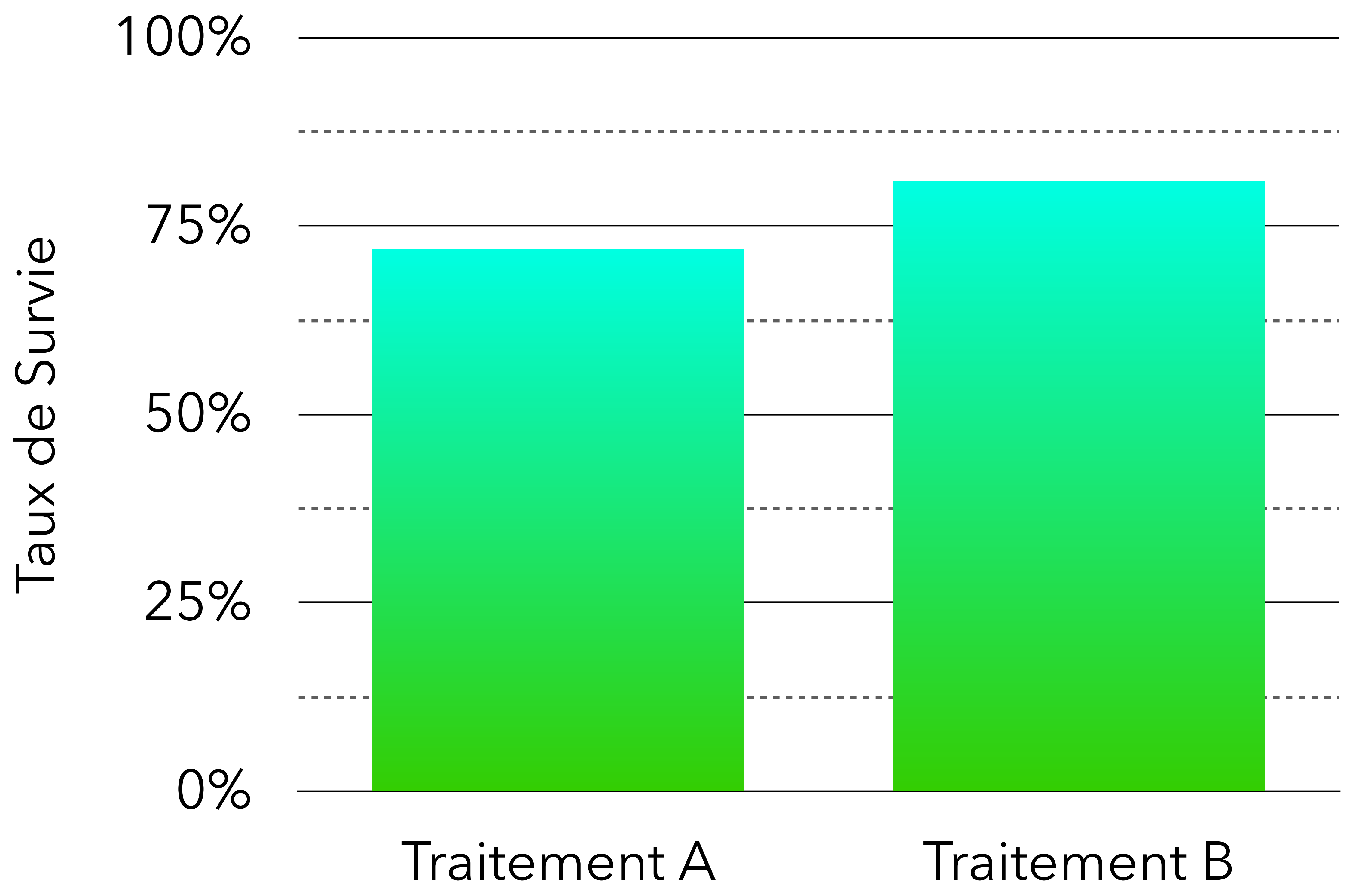
$$F_{Score} = 2\frac{Pre * Rec}{Pre + Rec}$$

Precision	**Recall**	**F Score**
0	0	0
1	1	1
0.5	0.4	0.444
0.7	0.1	0.175
0.02	1	0.0392

Paradoxe de Simpson

Contexte

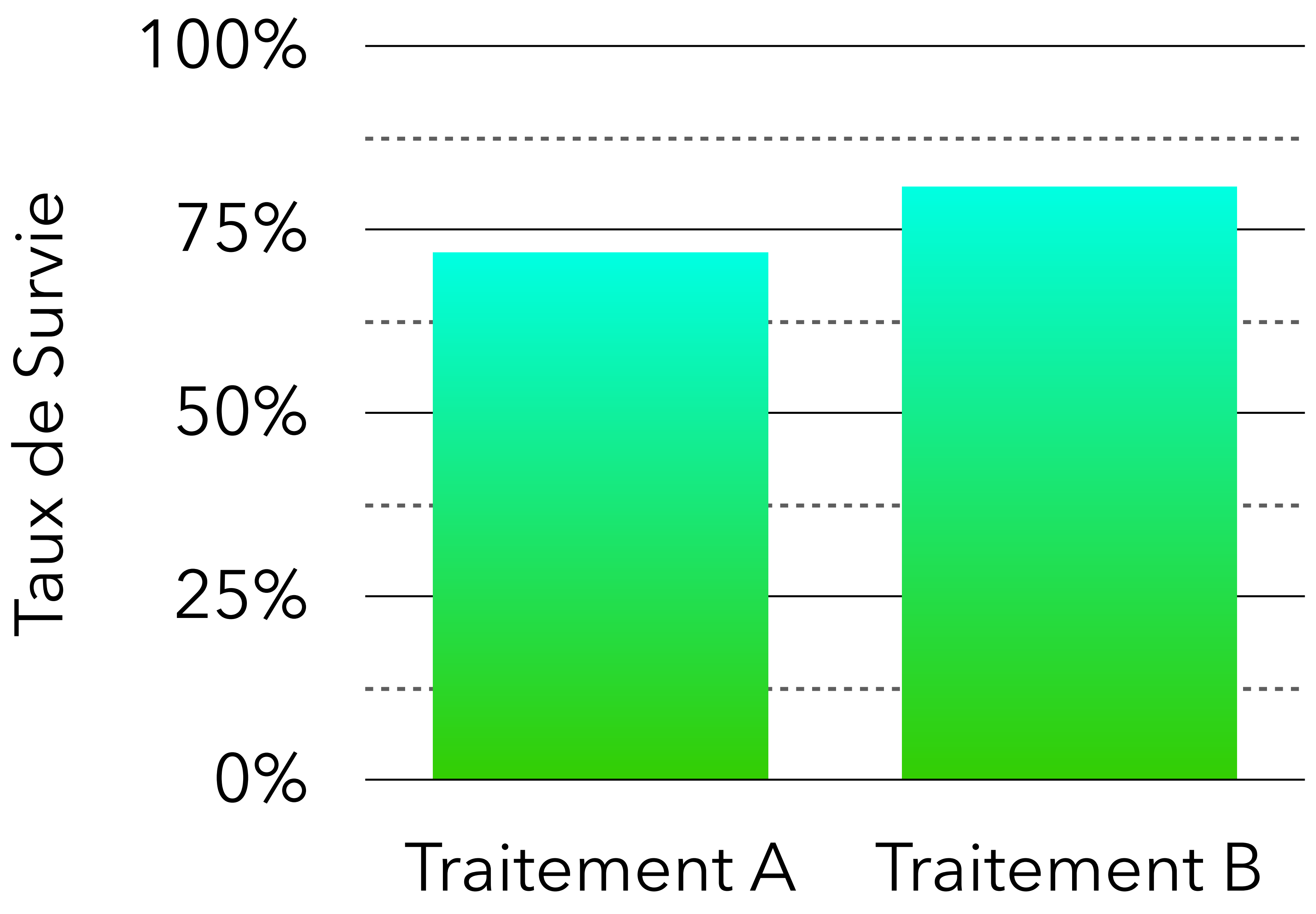
- Deux traitements sont testés pour des caillots sanguins dans les reins.
- Le traitement A est donné au groupe A, et le traitement B au groupe B.
- Les deux groupes sont de tailles égales.
- On observe les résultats suivants:

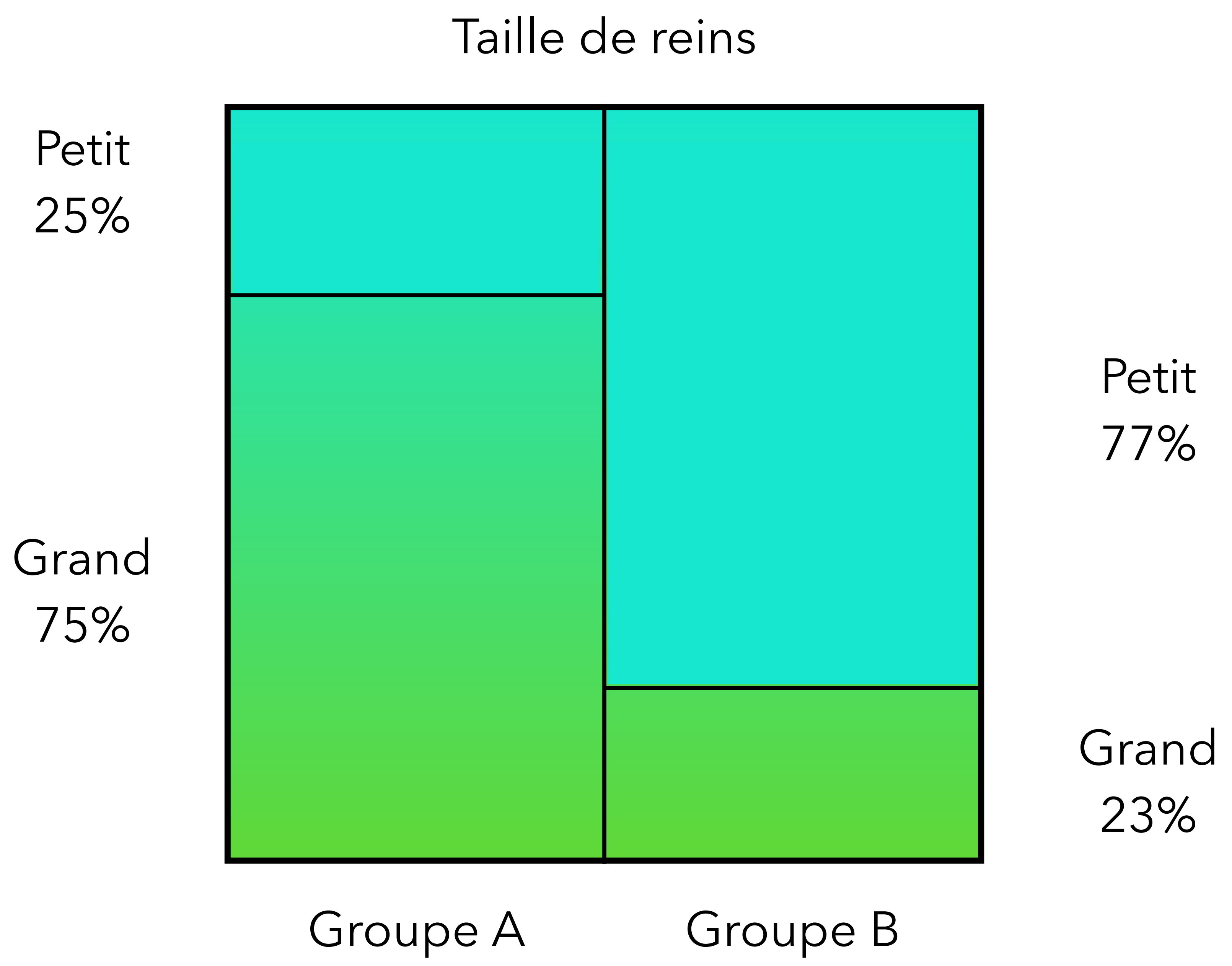


Paradoxe

Paradoxe : Le traitement A est en fait plus efficace que le traitement B !

- On trouve plus tard qu'en fait la <u>taille du rein</u> est un facteur important de survie pour cette maladie et que les groupes n'étaient pas répartis équitablement sur cette variable.

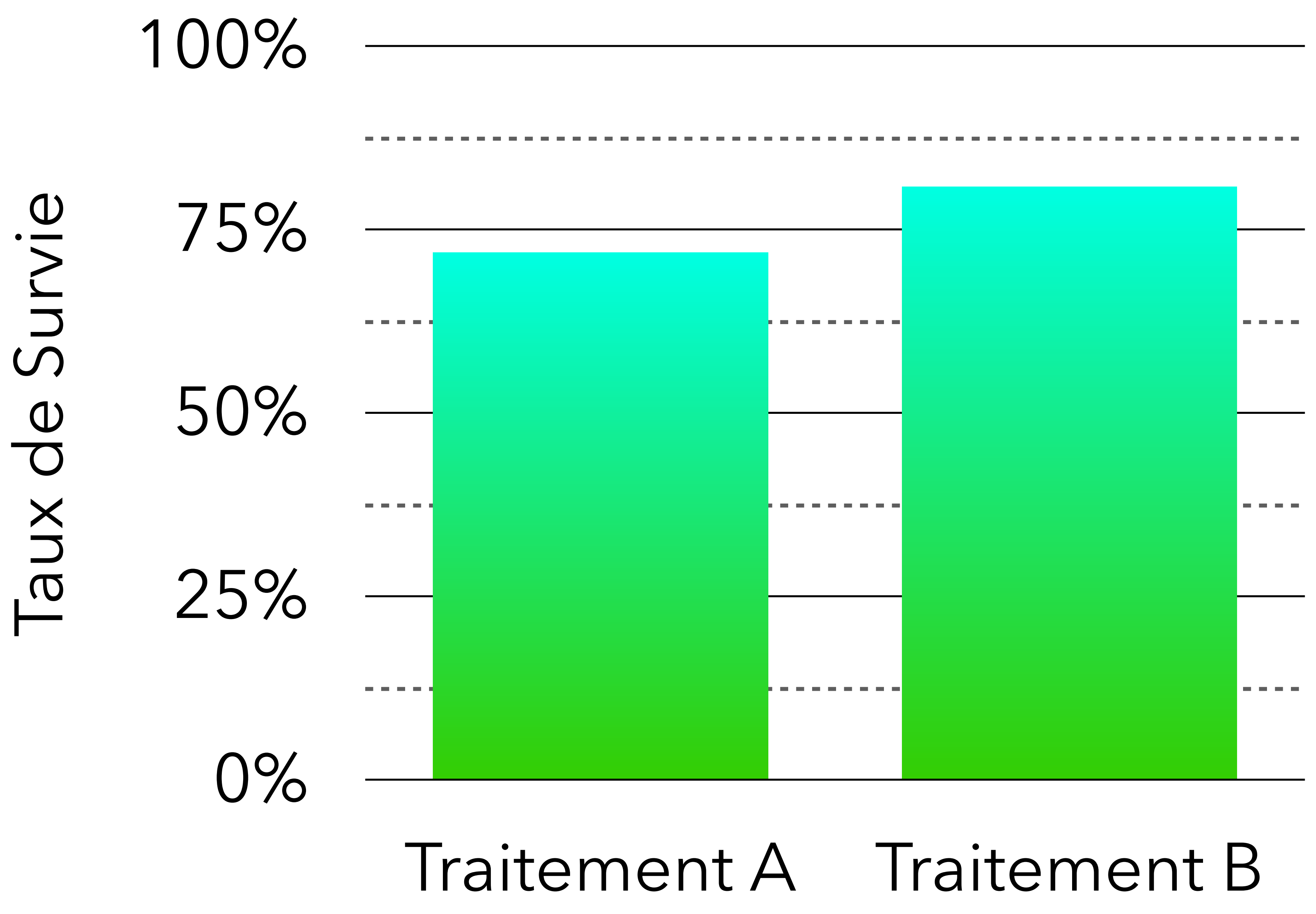


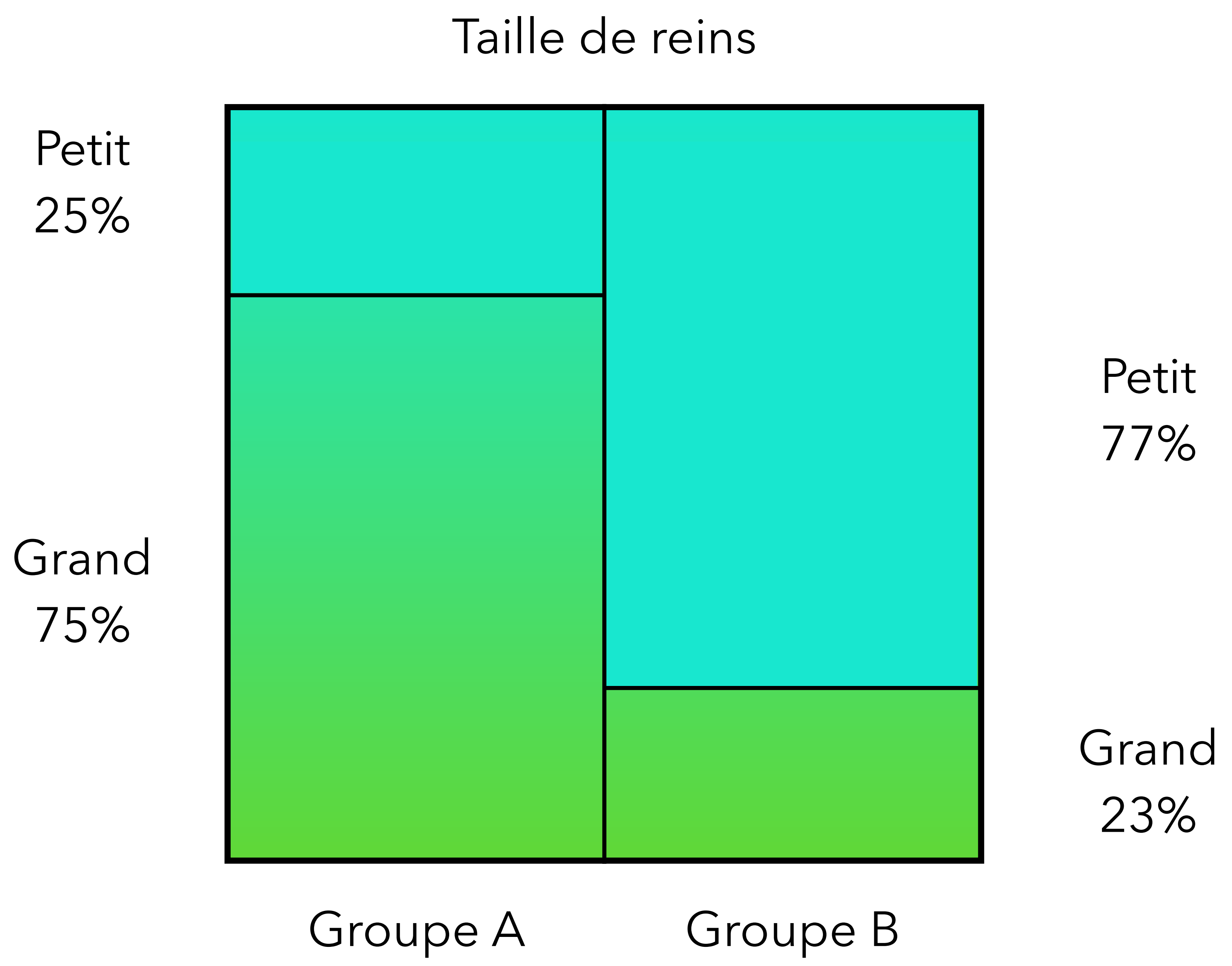


Résolution

Les patients avec un petit rein ont plus de chances de survie que les patients avec un grand rein.

➡ Du à la répartition des groupes, les patients du groupe B avait une meilleure chance de survie des le départ!





Parce qu'on ne connait pas toujours tous les facteurs influençant les résultats d'un test, avoir une population suffisantes qu'on répartie aléatoirement dans des groupes est primordiale.