

## Prediction Model



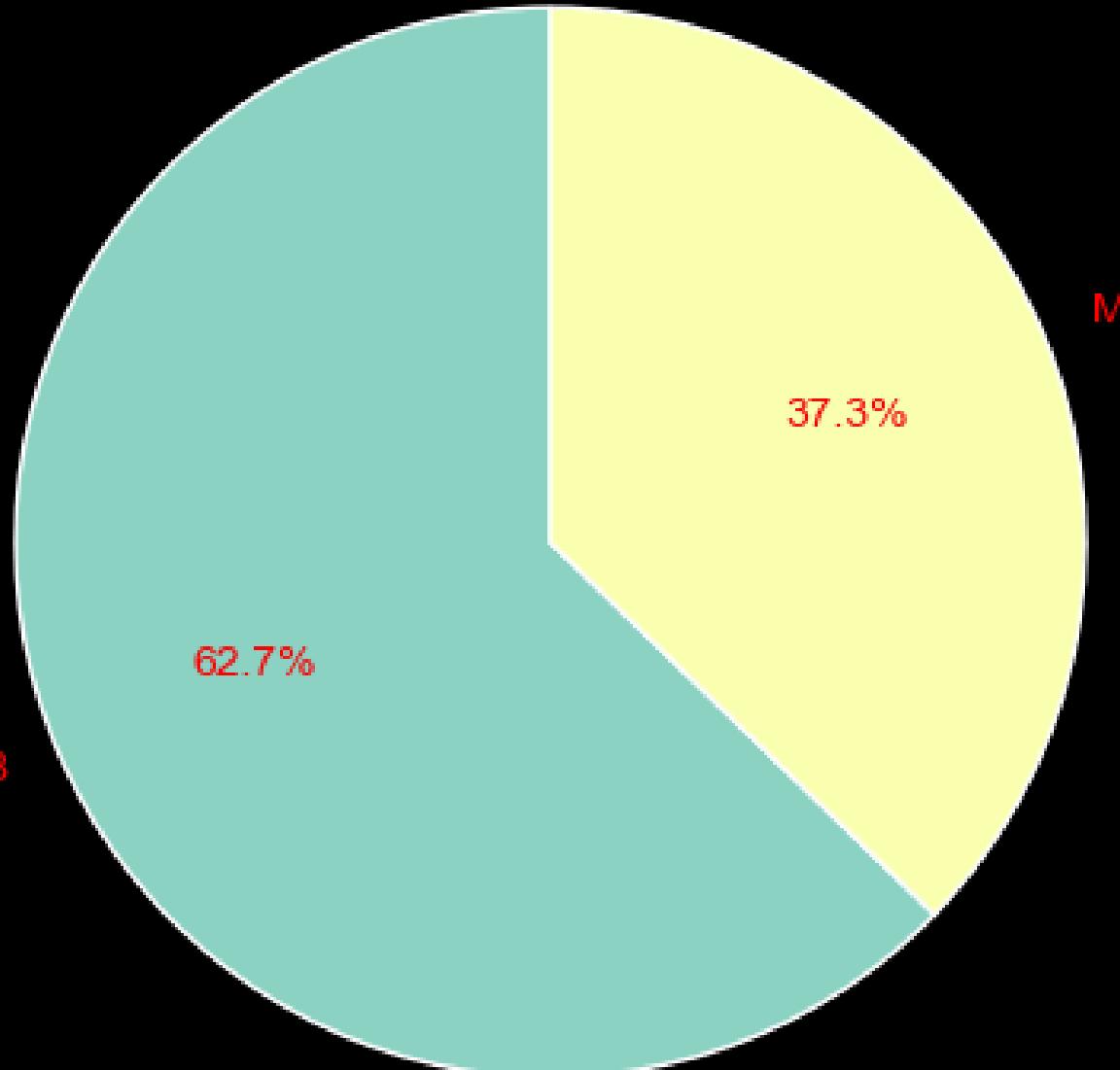
# Prédiction du cancer du sein

Utilisation de la régression logistique pour prédire la malignité des tumeurs à partir de caractéristiques morphologiques

569 observations

Régression Logistique

Distribution des classes (B vs M)



## Distributions des classes

Sur 569 tumeurs analysées, la majorité sont bénignes. Cette distribution reflète la prévalence naturelle des tumeurs du sein.

# Familles de Variables Morphologiques

**\_se**

Erreur standard mesurant la variabilité des observations

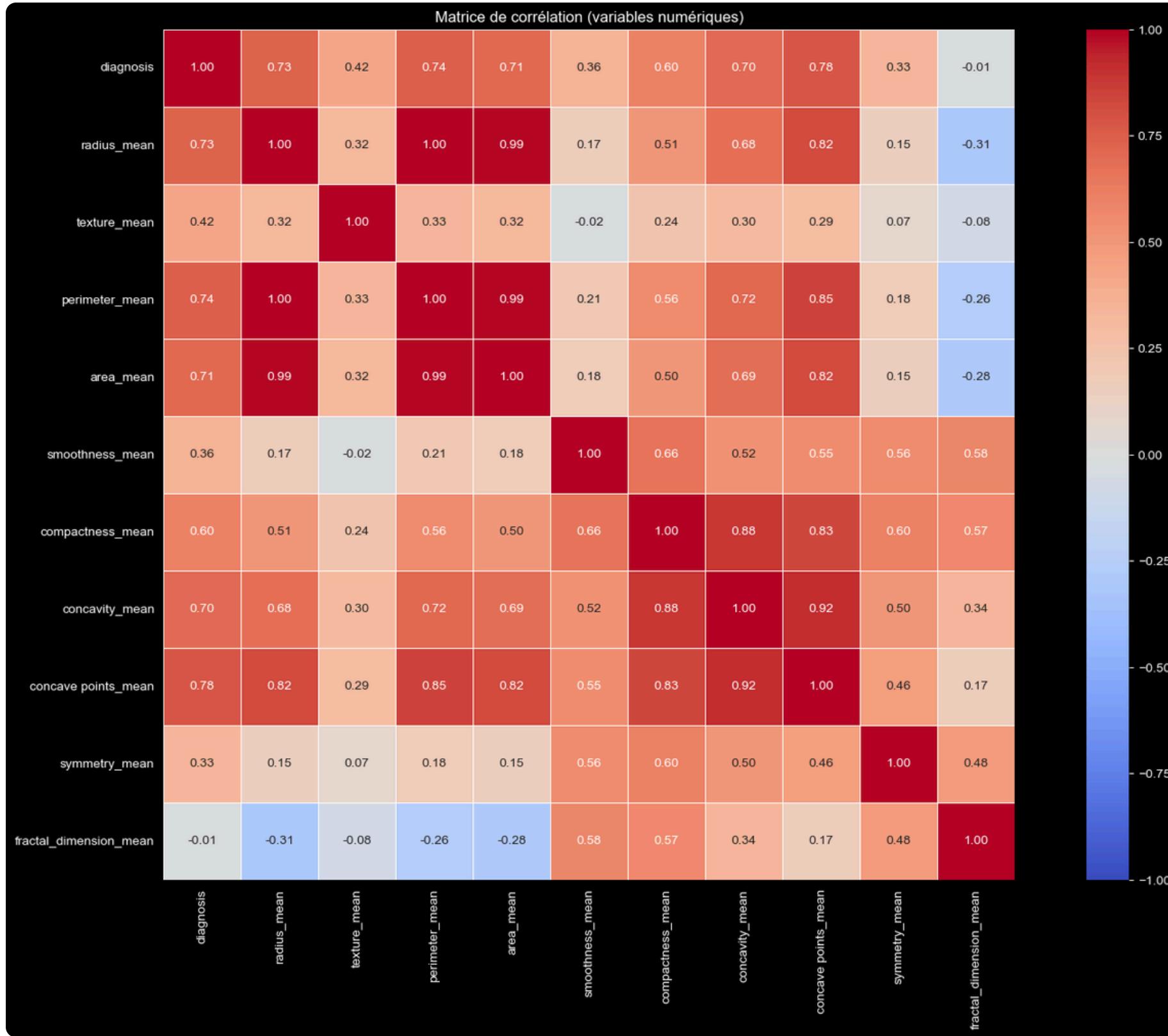
**\_mean**

Valeur moyenne des caractéristiques morphologiques de la tumeur

**\_worst**

Valeur la plus extrême observée pour chaque caractéristique

# Matrice de corrélation



La heatmap révèle des corrélations fortes entre certaines caractéristiques morphologiques et le diagnostic.

Les variables liées à la forme et à la taille montrent les associations les plus significatives avec la malignité.

# Variables sélectionnées

## **perimetre\_mean**

Mesure la taille globale de la tumeur, indicateur clé de la croissance cellulaire

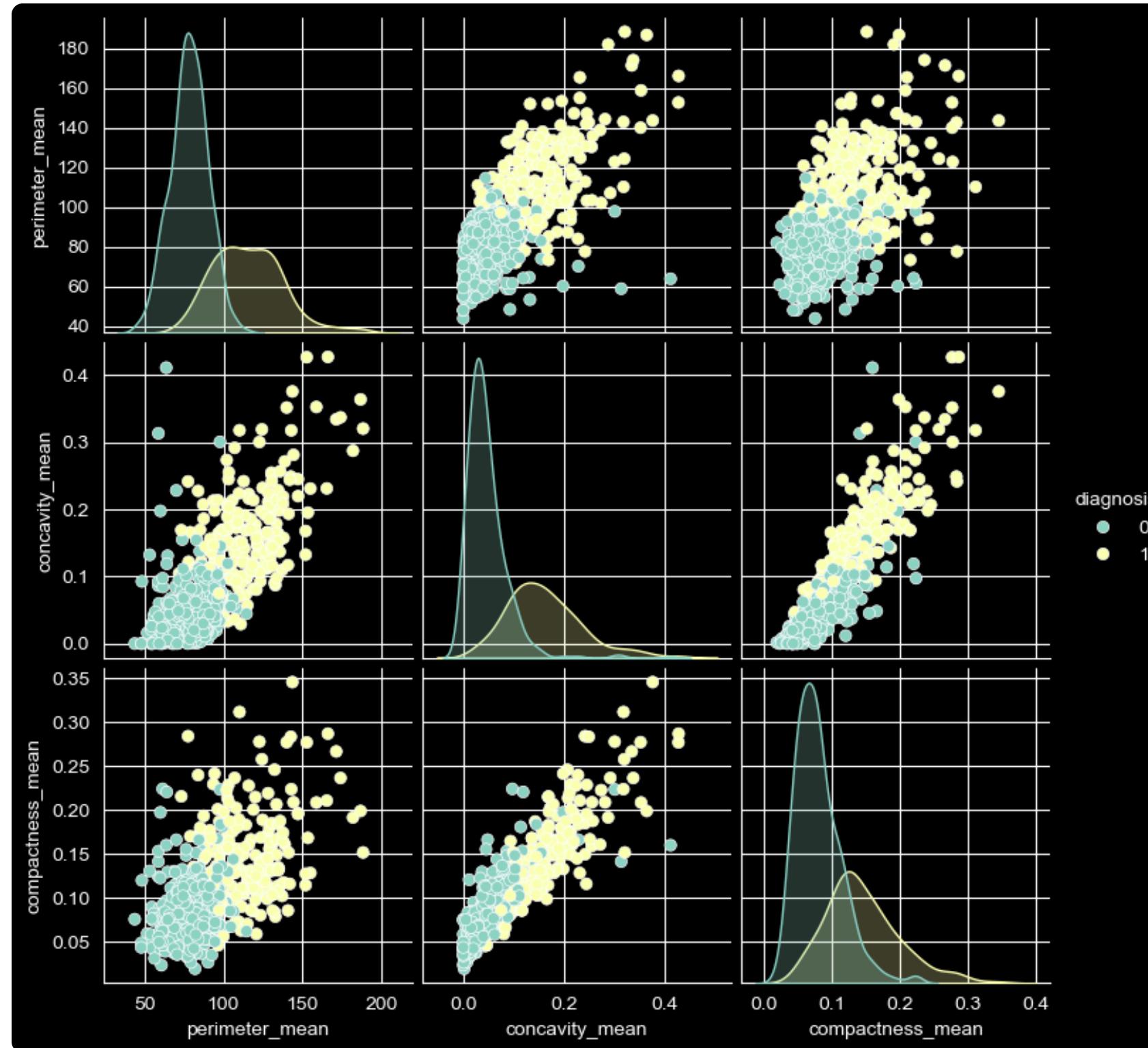
## **concavity\_mean**

Mesure le degré d'irrégularité du contour, révélateur de l'agressivité tumorale

## **compactness\_mean**

Mesure la densité et la concentration de la forme, indicateur de structure cellulaire

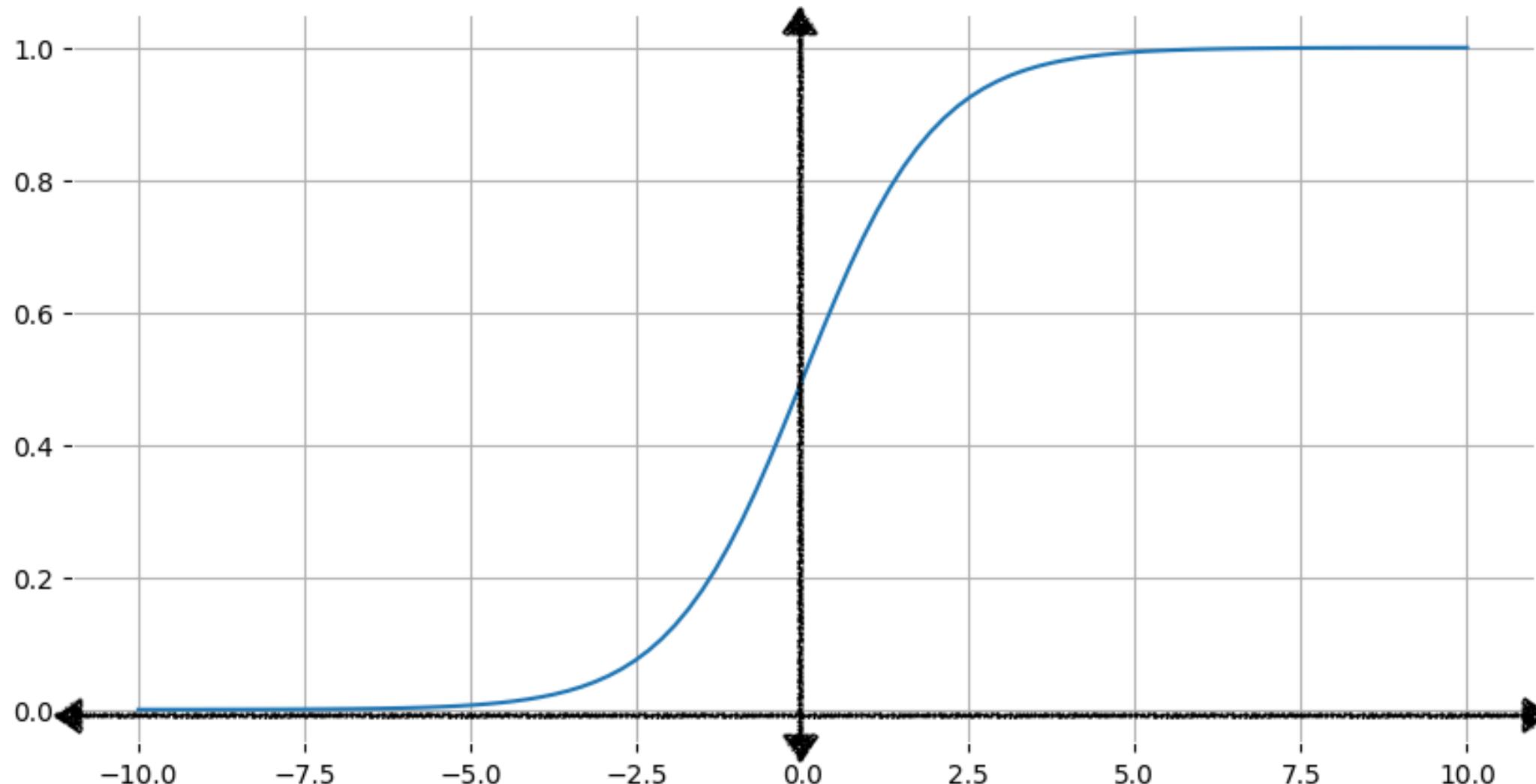
# Variables sélectionnées



Les graphiques montrent une séparation claire entre tumeurs bénignes (bleues) et malignes (jaunes).

Les tumeurs malignes présentent systématiquement des valeurs plus élevées pour les trois variables sélectionnées.

# Régression logistique



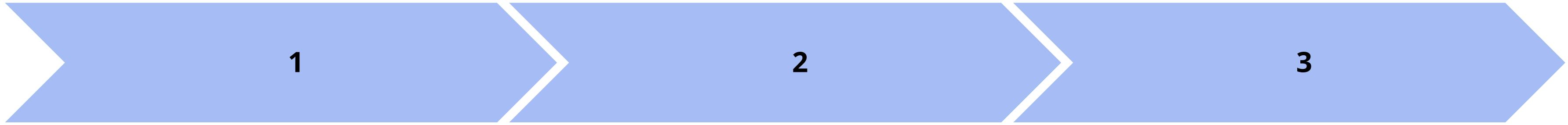
## Fonction Sigmoïde

La courbe en S transforme les valeurs continues en probabilités entre 0 et 1.

Cette fonction permet d'estimer le risque de malignité de manière progressive et interprétable.

# Performance des modèles

Pour une observation donnée



**perimeter\_mean**

Probabilité estimée : ~8%

**concavity\_mean**

Probabilité estimée : ~30%

**compactness\_mean**

Probabilité estimée : ~37%

# Modèle multivarié



**Variables combinée**

Périmètre + Concavité + Compacité

**Probabilité estimée**

~9% pour l'observation étudiée

# Conclusion

## Caractéristiques morphologiques

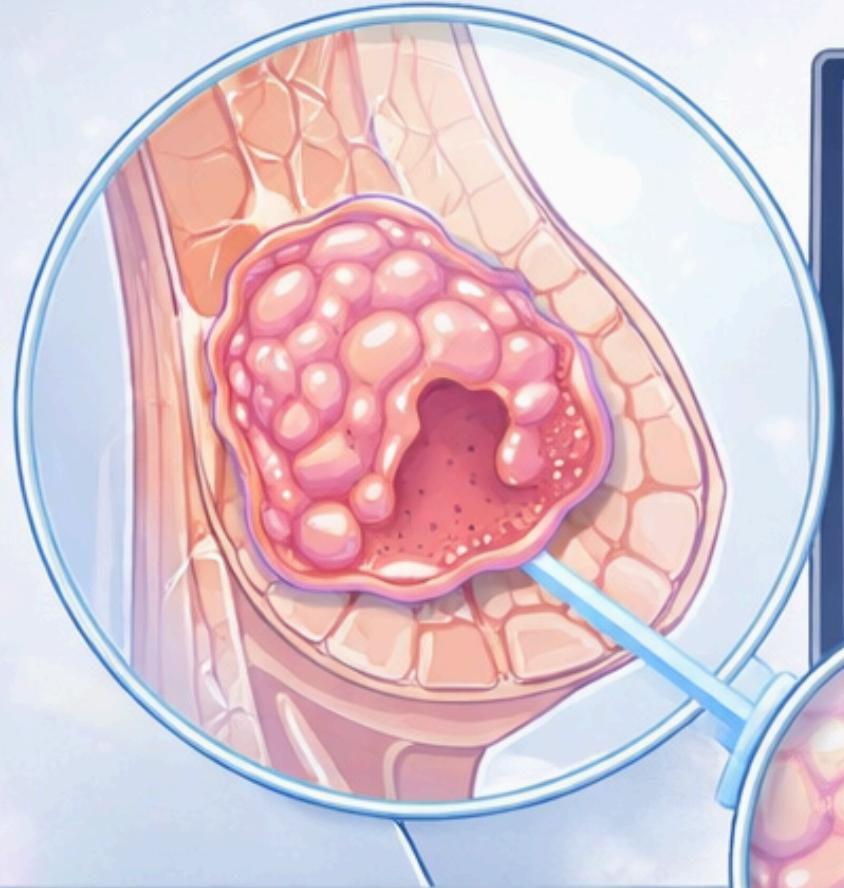
L'irrégularité de la forme est fortement associée à la malignité des tumeurs

## Estimation du risque

La régression logistique permet d'estimer une probabilité de manière cohérente et fiable

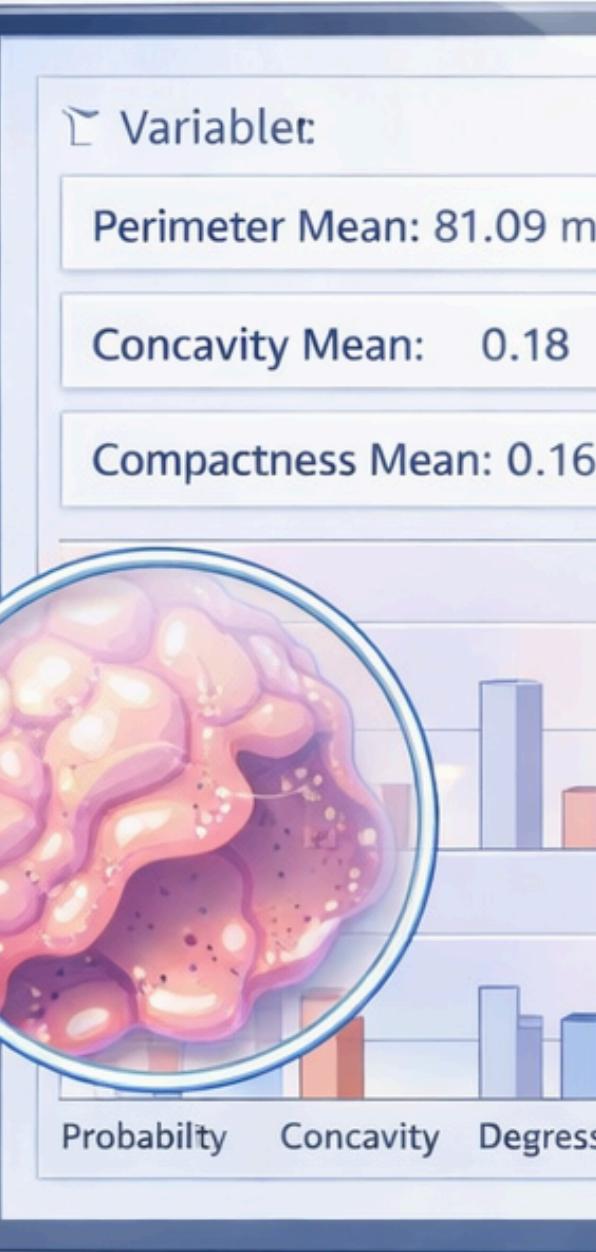
## Interprétabilité

Le modèle reste transparent et adapté à une analyse explicative en contexte médical



### Morphological Irregularities

Irregular contours & concavity associated with malignancy.



### Estimate Risk

Provides a probability that a tumor is malignant



### Interpretation

Identifies and quantifies predictors of malignancy