

PRESENTATION

Prédire un revenu

SOMMAIRE

MISSION

PRESENTATION DES DONNEES

DESCRIPTION

CREATION DES REVENUS PARENTS

PREDICTION DE REVENU ENFANT

MISSION

OBJECTIF	Prédire un revenu enfant
MISSION	Modéliser selon les pays
CONDITION	Revenus simulés des parents
DONNEES	revenus : world income Gini : banque mondiale

PRESENTATION DES DONNEES

PAYS (Banque Mondiale)

Année utilisée 2010

Nombre de pays 81 sur 160 (193)

Population couverte 79 %

REVENUS DES PAYS (World Income Distribution)

percentiles 100 quantiles

intérêt rendre les pays comparables

PRESENTATION DES DONNEES

Problématique

Données manquantes pour le Gini

Méthode de résolution

**trouver l'année la plus remplie
remplir avec la moyenne du pays**

Explication

La variation du Gini faible sur 20 ans

PRESENTATION DES DONNEES

Problématique

(fichier world income distribution)

Données numériques pour les quantiles

Méthode de résolution

méthode de classe `pd.read_csv`

signalement : virgule=décimale

DESCRIPTION

5 Gini les plus hauts

**Afrique du Sud(63,55), Namibie, Botswana,
Brésil, Honduras**

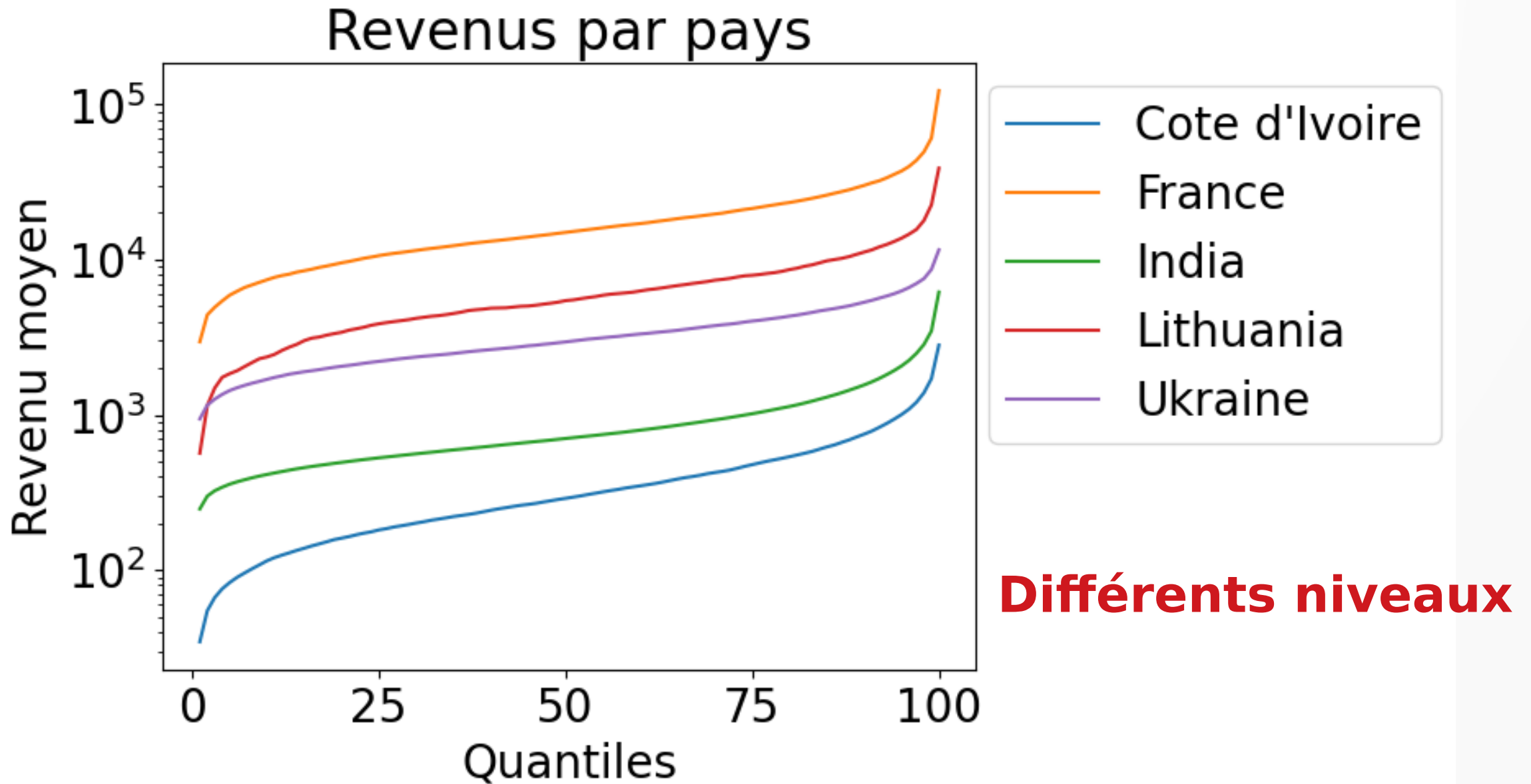
5 Gini les plus bas

**Danemark, République de Slovaquie,
Ukraine, République tchèque,
Slovénie (24,89)**

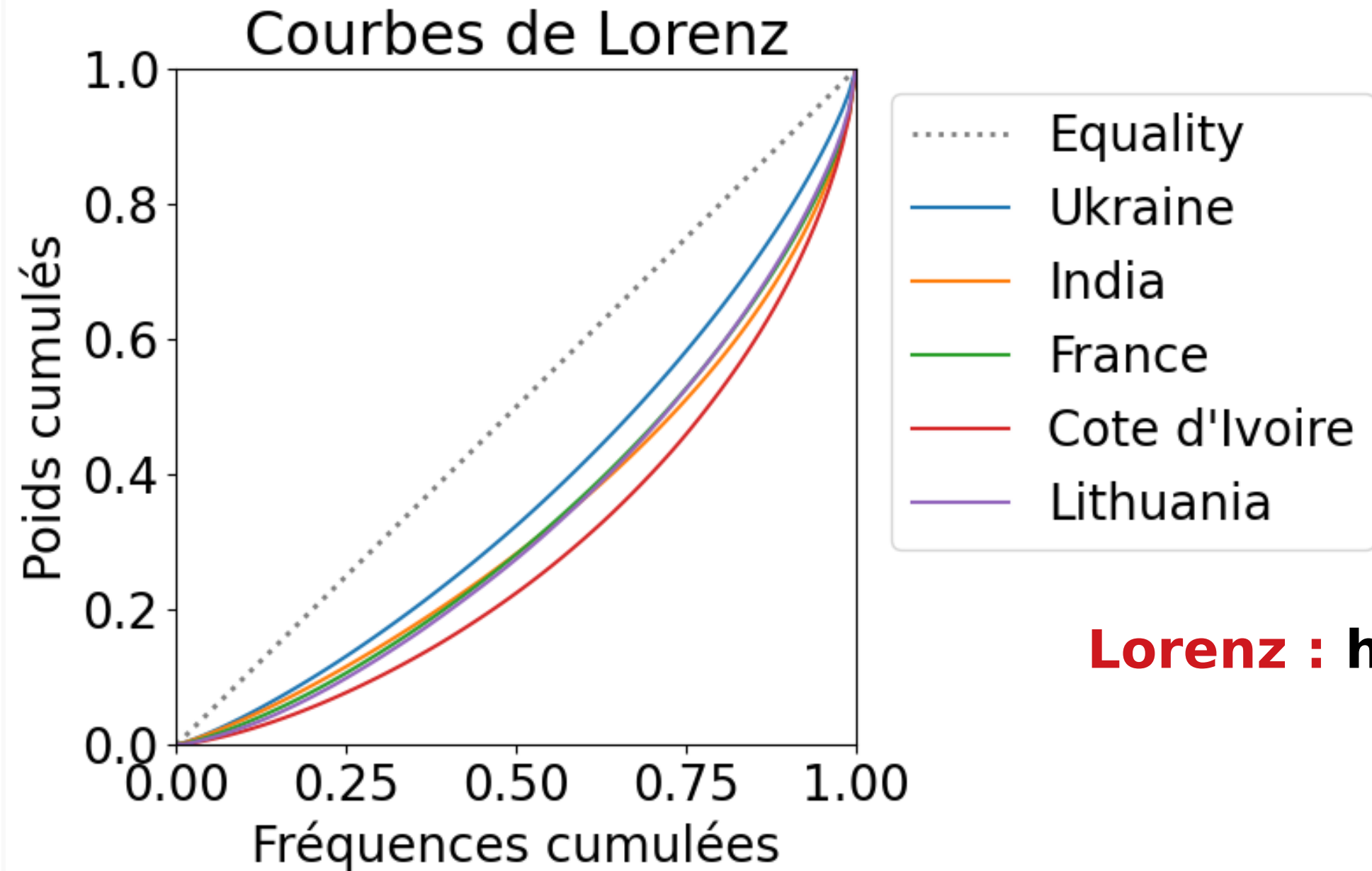
Gini moyen : 38,89

Place de la France : 129

DESCRIPTION

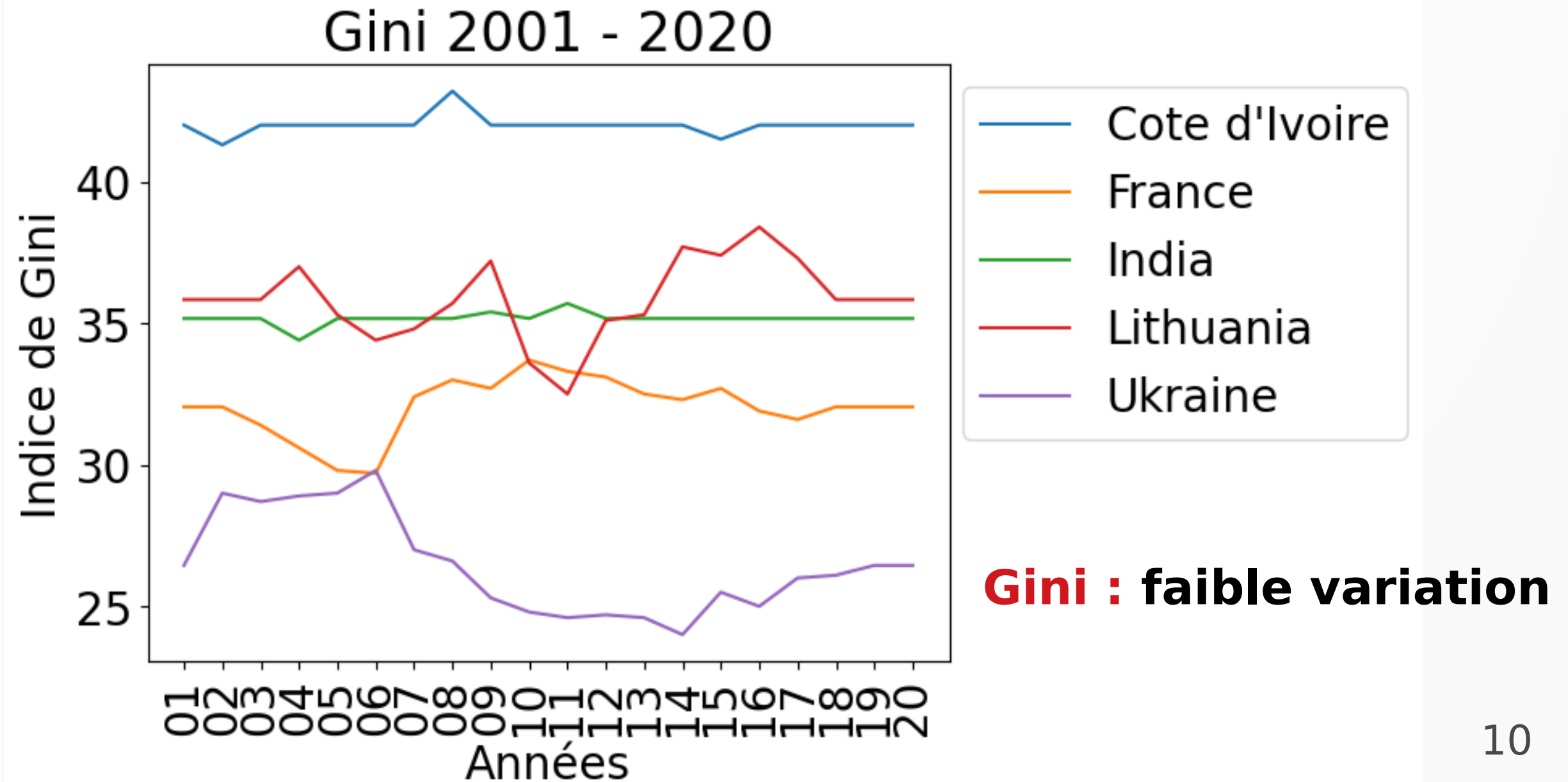


DESCRIPTION



Lorenz : homogénéité

DESCRIPTION



CREATION DES REVENUS PARENTS

Créer les coefficients d'élasticité

- **fichier GDIM : données manquantes**
- **imputation par médiane de région**
- **50 % de données manquantes**

Générer les probabilités conditionnelles

- **Générer les classes parents (loi normale)**
- **Calculer les cardinaux**
- **Diviser les classes parents par les cardinaux**

Générer les clones

- **Multiplier les probabilités par les clones**

CREATION DES REVENUS PARENTS

Problématique

Temps d'exécution du script très long +10min

Méthode de résolution

Ecriture d'un code optimisé et commenté

Evaluer la performance avec un chronomètre

Explication

méthode de calcul scientifique : Numpy

utilisation des méthodes Pandas

CREATION DES REVENUS PARENTS

Problématique

**Impossible de cloner 500 individus
décimales des probabilités conditionnelles**

Méthode de résolution

Générer 1000 individus

Synthèse

- de + 10min à 15 secondes**
- générer 2 fois plus d'individus**

PREDICTION DE REVENU ENFANT

Vérification avant régression

Normalité des revenus

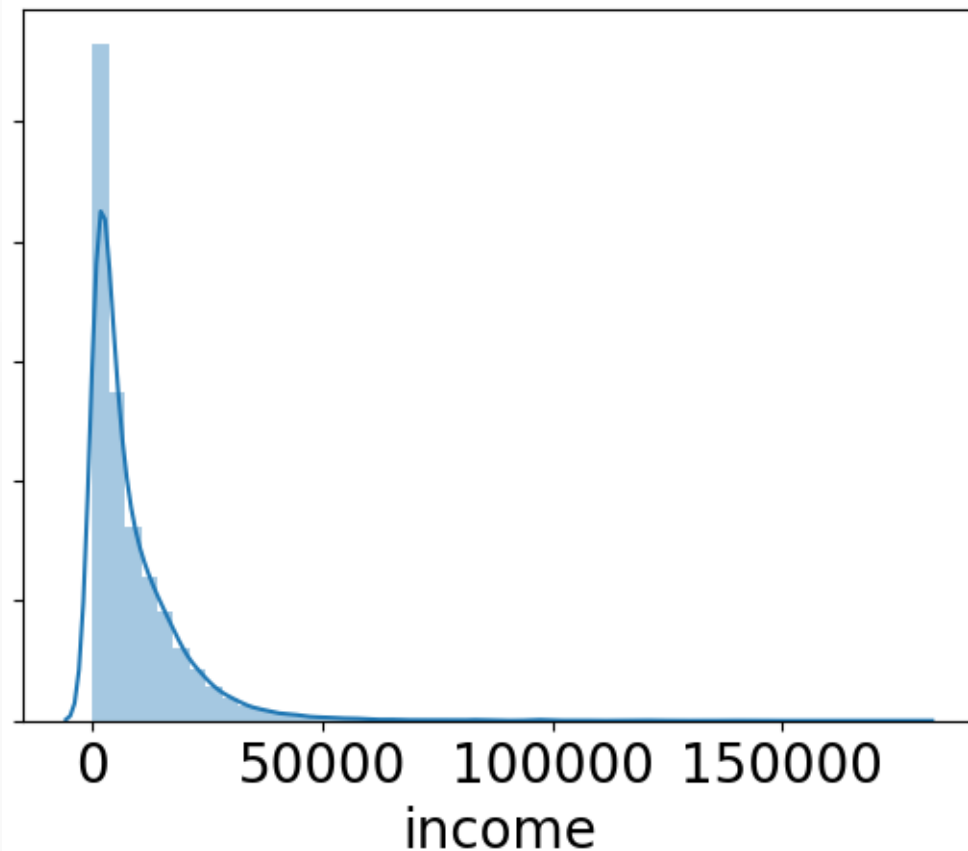
Formes des revenus

- **sans transformation**
- **avec transformation logarithme**

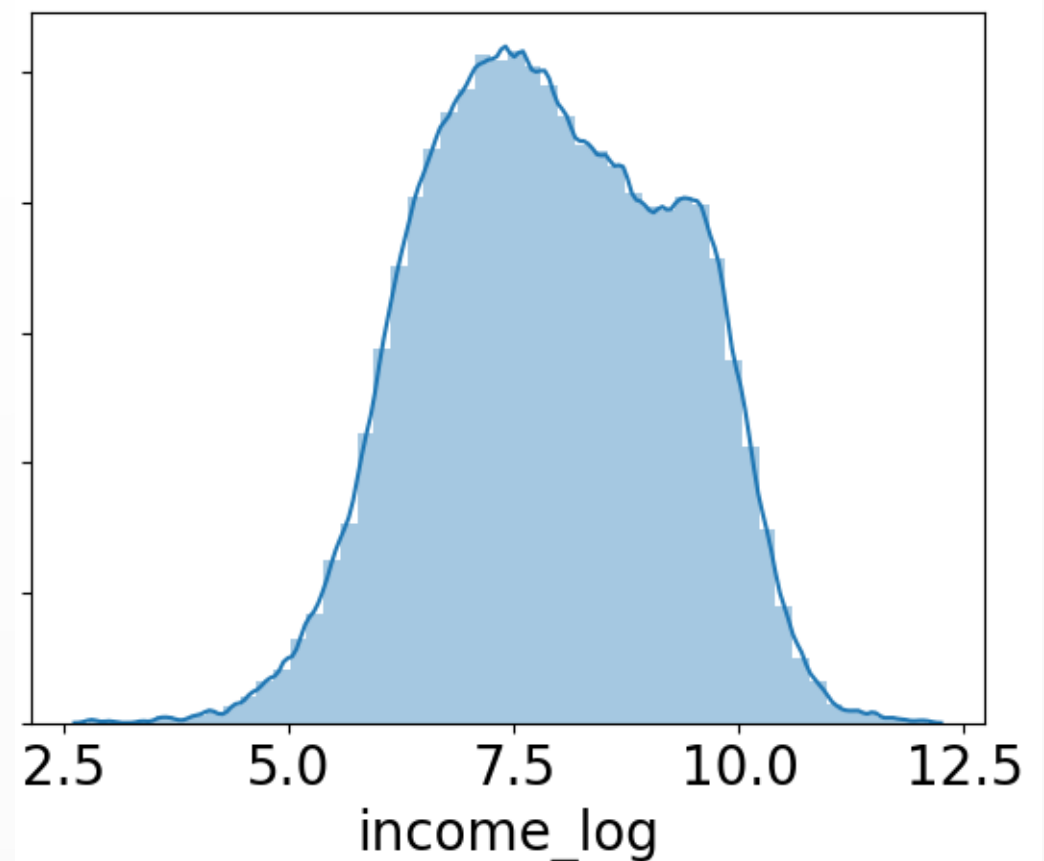
Explication

« On devient ce que l'on mange »

PREDICTION DE REVENU ENFANT



Sans transformation



Avec transformation

PREDICTION DE REVENU ENFANT

ANOVA

Variable à expliquer : logarithme des revenus

Variable explicative : Pays

Méthode des moindres carrés

Variance expliquée 72 %

PREDICTION DE REVENU ENFANT

Problématique

Améliorer la qualité du modèle

Méthode de résolution

Backward Elimination

- **fixer un seuil de pvalue**
- **inclure tous les pays**
- **enlever le pays où la pvalue supérieur seuil**
- **réitérer la regression tant qu'il reste un pays**

PREDICTION DE REVENU ENFANT

Problématique

Boucle infinie pour le backward elimination

Méthode de résolution

- **trouver la raison**
- **filtrer les pays déjà enlevés**

Synthèse

- **78 % : gain de 6 %**
- **enlever les pays hors distribution**

PREDICTION DE REVENU ENFANT

Régression linéaire multiple 1

Variable à expliquer : logarithme des revenus

Variables explicatives :

- **revenu moyen**
- **indice de Gini**

Variance expliquée 78 %

Coefficients

- **revenu moyen : 0,99**
- **indice de Gini : - 0,02 (impact négligeable)**
- **constante : 0,52**

PREDICTION DE REVENU ENFANT

Régression linéaire multiple 2

Variables explicatives :

- **revenu moyen + indice de Gini du pays**
- **classe des revenus des parents**

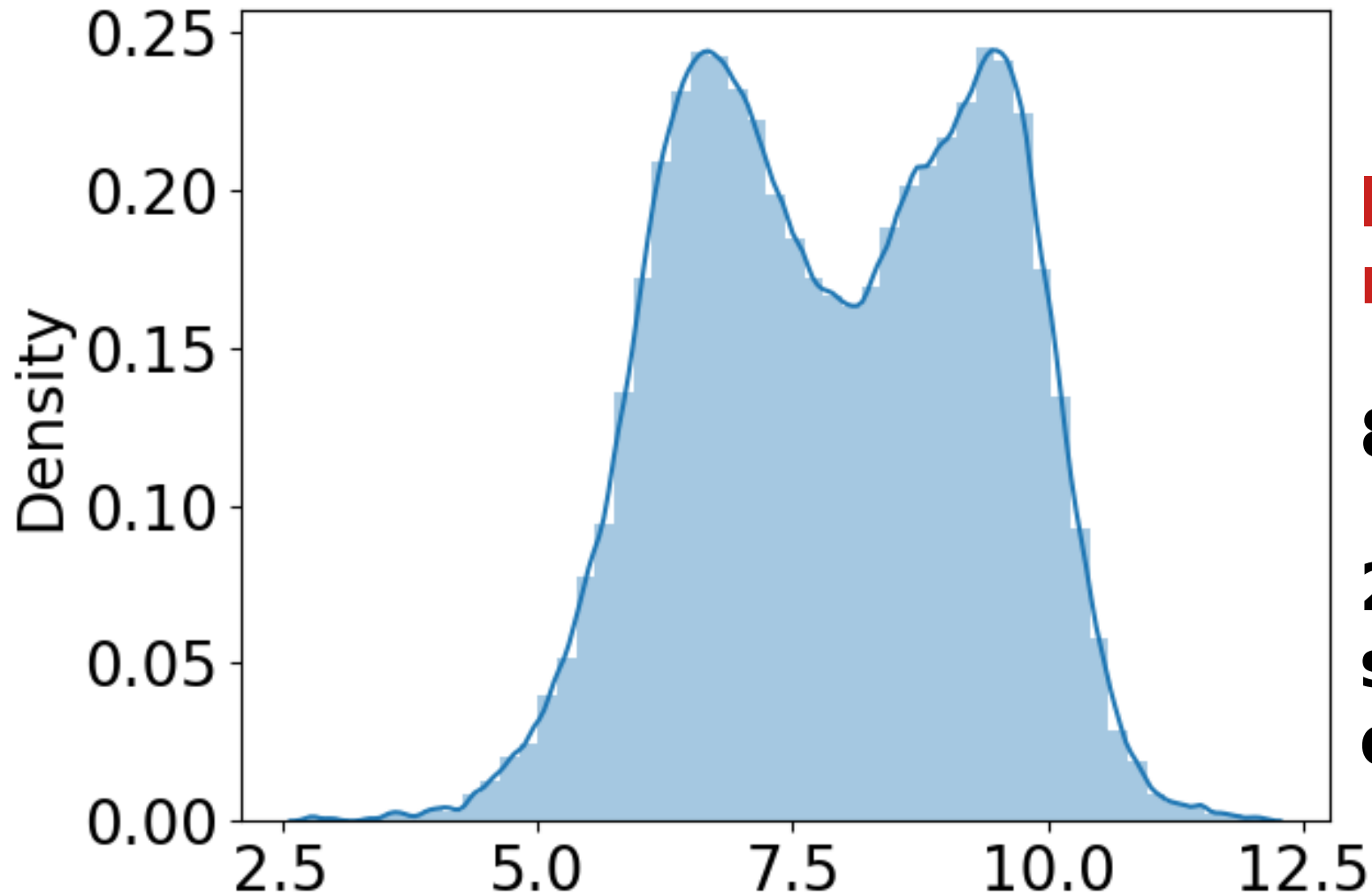
Variance expliquée 82 %

Coefficients

- **revenu moyen : 0,99**
- **indice de Gini : - 0,02 (impact négligeable)**
- **classe des parents : 0,01**
- **constante : - 0,02**

PREDICTION DE REVENU ENFANT

modèle imprécis : pourquoi ?



**logarithme des
revenus**

80 % population

**2 populations
statistiques
différentes**

CONCLUSION

IMPACTANT

**MOYENNE DU PAYS
CONNAISSANCE REVENUS PARENTS**

NON IMPACTANT

INEGALITES DE REVENUS

**ELEMENT CLEF
2 CATEGORIES DE POPULATIONS**

QUESTIONS
