

# ETUDE DE MARCHÉ: DU POULET POUR TOUS

---

SOUTENANCE DU PROJET 5

PARCOURS DATA ANALYSTE VERSION I

TRAORE LASSINA JAMAL

# PLAN

---



# INTRODUCTION

## ENJEUX

---

- Vendre plus de poulet en exportant
- L'aviculture peut réduire la sous-nutrition à travers ses multiples produits dérivés
- Le poulet représente la deuxième viande la plus consommée au monde
- Le poulet fait partie intégrante de tous les régimes alimentaires
- Une bonne maîtrise du marché avicole permet une bonne couverture en protéine animale
- Ainsi , connaître la demande réelle du marché avicole passe par
  - L'évolution démographique
  - Les régimes alimentaires des pays

# INTRODUCTION ENJEUX

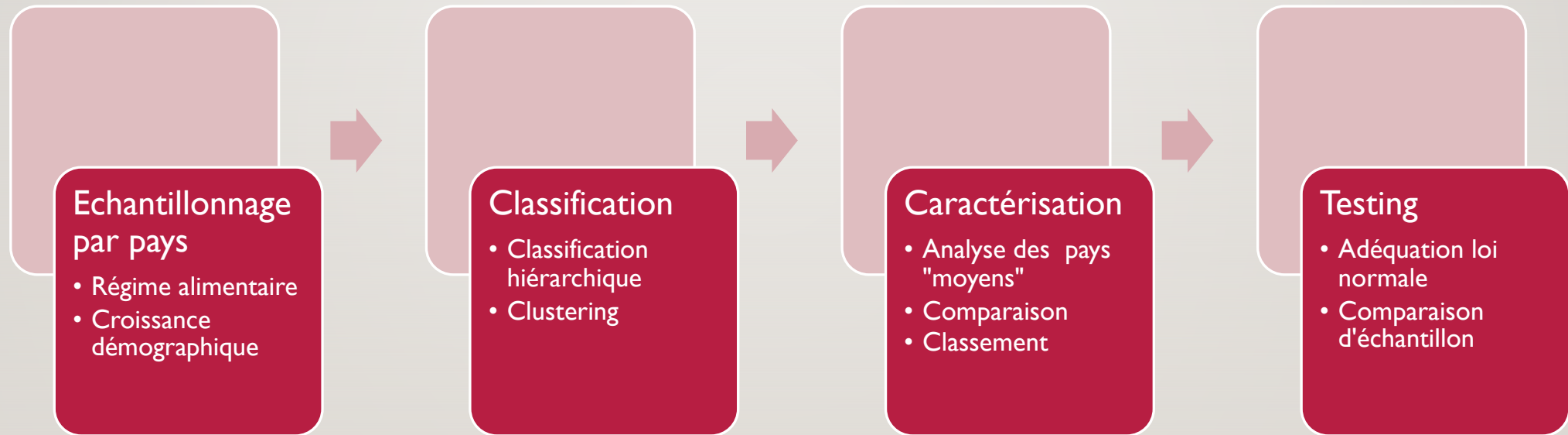
---

- Mission d'étudier le Marché en ciblant des pays
- Ces pays doivent assurer une bonne rentabilité
- En soutenant la croissance de notre chiffre d'affaire



# DÉMARCHE

---



DÉMARCHE



# DÉMARCHE ÉCHANTILLONNAGE

---

- Operations sur Dataframes
- Collecte des données sur le site de la Food and agricultural organisation (FAO)
- Creation de dataframe et opérations diverses
  - Compilation *dfpopulations* (2013 et 2018)
  - Calcul de ratio entre les années pour estimer la croissance démographique
  - Restriction et filtrage de colonnes d'animaux et végétaux 2018 (tables *animaux* et *végétaux*)
  - Prétraitement sur les deux tables (suppression de Valeurs aberrantes)
  - Fusion de dataframes (Population, animaux )

# ECHANTILLONNAGE

## TABLE D'ÉTUDE

---

	Zone	ratioprot	ratiochr	disponibilite par habitant prot(kg/an/hab)	Disponibilité alimentaire (Kcal/personne/jour)
0	Afghanistan	19.441441	21.667976	20.25750	2038.0
1	Afrique du Sud	42.716605	9.506215	30.79505	2895.0
2	Albanie	53.375983	-9.139616	42.21955	3361.0
3	Algérie	26.922239	7.702510	33.51430	3321.0
4	Allemagne	60.893961	0.479892	38.37975	3554.0

\*ratioprot= taux de protéine animale sur la disponibilité alimentaire

\*ratiochr= évolution de la croissance démographique sur des années



# DEMARCHE- RESULTATS



# CLASSIFICATION HIÉRARCHIQUE CLUSTERING

---

	Zone	ratioprot	ratiochr	disponibilite par habitant prot(kg/an/hab)	Disponibilité alimentaire (Kcal/personne/jour)	cluster
0	Afghanistan	19.441441	21.667976	20.25750	2038.0	2
1	Afrique du Sud	42.716605	9.506215	30.79505	2895.0	4
2	Albanie	53.375983	-9.139616	42.21955	3361.0	3
3	Algérie	26.922239	7.702510	33.51430	3321.0	4
4	Allemagne	60.893961	0.479892	38.37975	3554.0	3

- Le nombre d'individus un peu trop ne permettait pas de mieux qualifier
- 5 clusters ont permis de dégager des groupes presque homogènes

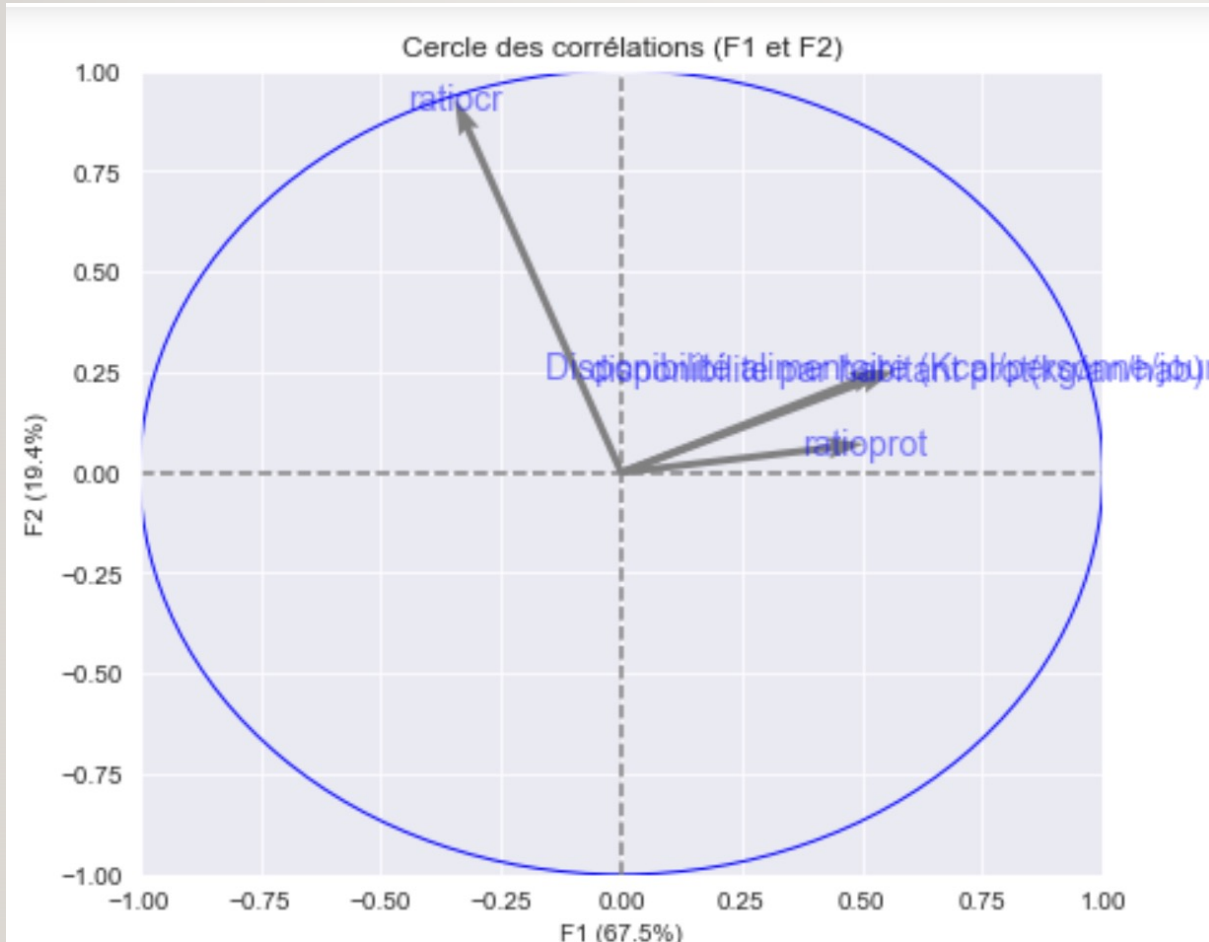
# ANALYSE D'ÉCHANTILLON

## ACP

---

- L'ACP nous a permis d'étudier :
  - La variabilité des individus (ressemblances et différences)
  - Trouver des variables synthétisant les variables initiales
  - La représentation à  $k$  dimension de sorte que  $k=2$  permet de créer un nuage de points
  - La projection des individus permet une meilleure appréhension de leur position sur le plan factoriel premier
  - Ainsi on peut voir les ressemblances des individus selon leur proximité
  - On y arrive grâce aux centrages et réductions des données pour les uniformiser et ne pas perdre les informations

# RESULTATS ACP



- ratiocr est très positivement corrélée à F2 quand dispo protéine par hab et ratioprot sont positivement corrélées à F1.
- Les deux var (dispo protéine par hab et dispo alim calorie) se confondent, donc elles sont très significantes pour F1.
- Les variables ratiocr et dispo de protéine forment parfaitement un angle droit donc on dirait qu'elles sont anti corrélées.



# INTERPRÉTATIONS

## CERCLE DE CORRÉLATION

---

- Les var (ratiocr et dispo alim en protéine animale) fortement corrélées à FI indiquent les consommations de viande des pays en lien avec les populations.
- Globalement tous les pays consomment de la viande et la var (ratioprot) fortement corrélée à F2 indique la croissance démographique des pays.
- Quand la démographie croît, la consommation de viande s'en trouve impactée puisque les disponibilités vont se modifier
- Ainsi les disponibilités croissent et le taux de protéine animale va augmenter du fait que la viande constitue une très grande part de consommation des foyers

## RESULTATS CLUSTERING- PROJECTION

---



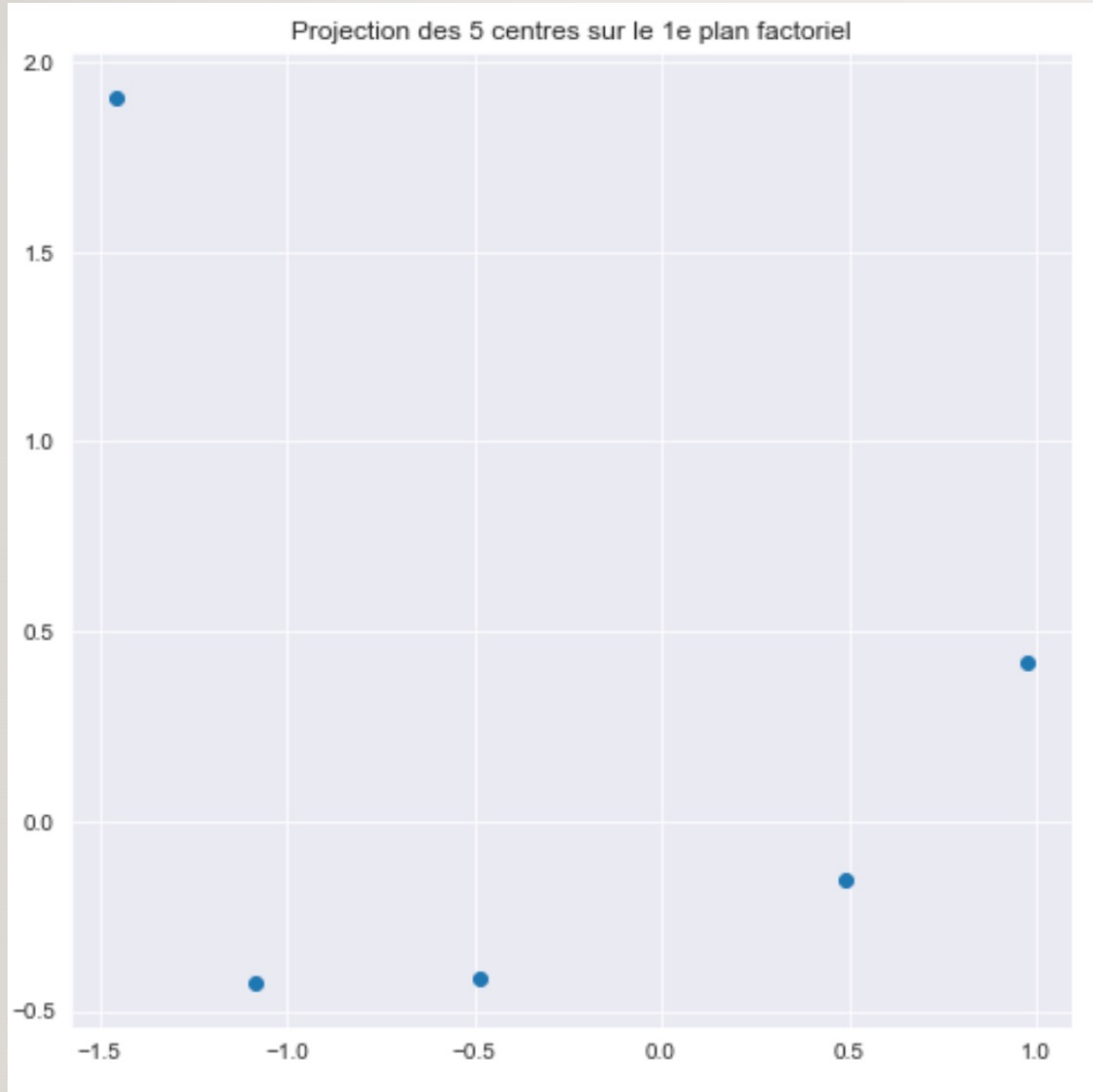
- Les 5 clusters sont visibles sur le premier plan factoriel
- 3 des clusters sont étalés sur F1 et 2 sur F2
- Le cluster en violet est remarquable par sa grande inertie interclasse et intraclasses
- Nous réduisons ainsi la dimensionnalité de notre échantillon

# RÉSULTATS

## TABLE CLUSTER ET VALEURS MOYENNES

---

	cluster	ratio prot	ratio cr	disponibilite par habitant prot(kg/an/hab)	Disponibilité alimentaire (Kcal/personne/jour)
		mean	mean	mean	mean
0	1	42.325646	38.664095	26.562267	2628.333333
1	2	25.271633	11.959363	22.445610	2433.875000
2	3	58.984880	1.547271	40.451844	3458.424242
3	4	35.911306	10.928050	34.297225	3171.937500
4	5	51.663433	2.582606	29.611831	2875.033898



## RÉSULTATS AFFICHAGE CENTROIDES

---

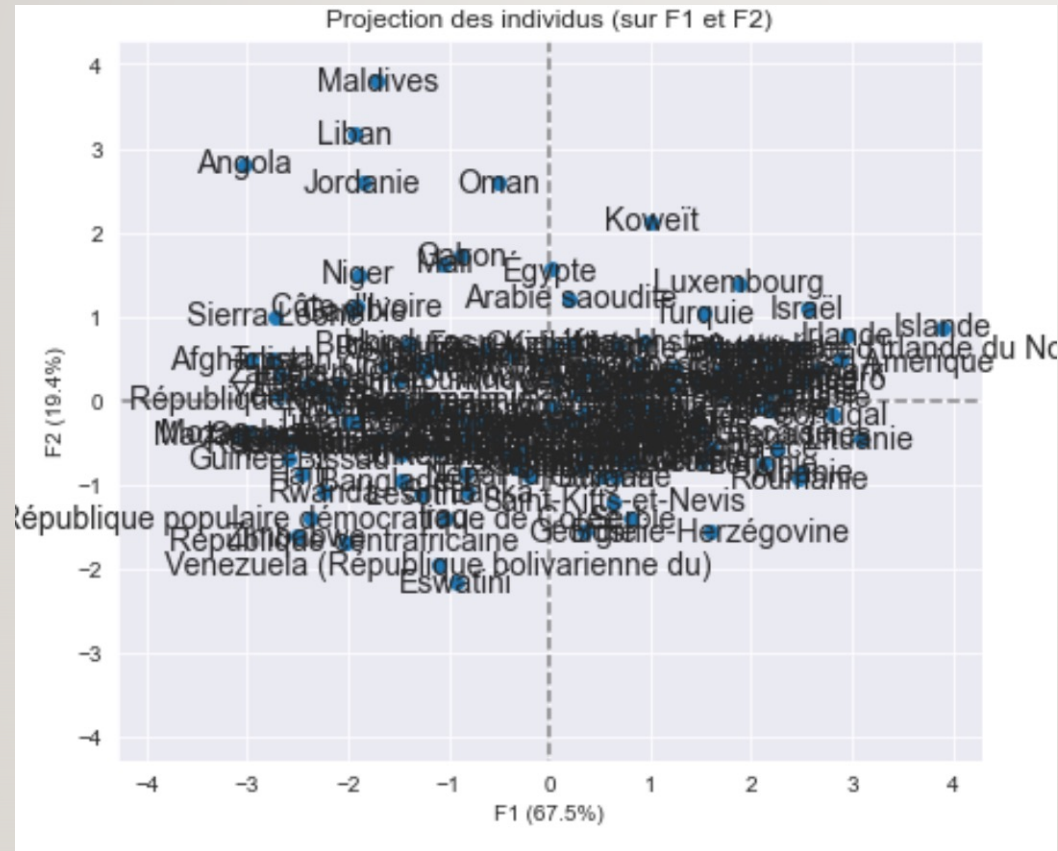
- 4 des centroids sont étalés sur F1 ayant des distances presque égales entre eux
- le cluster 1 est très corrélé avec F2 et se détache fortement des autres

# RÉSULTATS RÉDUCTION DE VARIABLES

- Le premier plan factoriel nous permet de réduire les dimensions de notre échantillon
- Le nuage de points réalisé nous permet de dégager deux principales variables :
- -ratio de croissance démographique
- -ration de protéine animale
- On peut voir une forte corrélation sur F1 qui est la consommation de viande
- Et une forte corrélation sur F2 représentant la croissance démographique



# RESULTATS ACP



## INTERPRÉTATION PROJECTION DES INDIVIDUS

- Sur FI , nous testons ce qui peut différencier l'Islande et l'Afghanistan et remarquons que l' Afghanistan a une population qui croit dix fois plus que celle de l'Islande mais le taux de protéine animale y est trois fois moins supérieure.

	Zone	ratio prot	ratio cr	disponibilite par habitant prot(kg/an/hab)	Disponibilité alimentaire (Kcal/personne/jour)	cluster	
68	Islande	71.668487	2.121212		53.4652	3656.0	3

```
tcp1=DFmerge.loc[DFmerge['Zone']=='Afghanistan']  
tcp1.head()
```

	Zone	ratio prot	ratio cr	disponibilite par habitant prot(kg/an/hab)	Disponibilité alimentaire (Kcal/personne/jour)	cluster	
0	Afghanistan	19.441441	21.667976		20.2575	2038.0	2

# INTERPRETATION

## ACP/COMPOSANTES PRINCIPALES

---

- On pourrait synthétiser F2 comme variable de croissance démographique et F1 comme la consommation de viande
- Globalement tous les pays mangent de la viande d'où un nuage de points touffu centré sur le plan factoriel mais la croissance démographique est fortement marquée chez certains pays comme ceux du cluster I ce qui indique la forte corrélation avec F2.
- On pourrait synthétiser nos 4 variables en 2 qui seront :
  - Consommation de viande (F1)
  - Croissance démographique (F2)

# CHOIX DU CLUSTER



# CARACTÉRISATION COMPARAISON ET CLASSEMENT

---

- Les centres des classes constituent les moyennes des observations par cluster
- Ces moyennes permettent de mieux qualifier les clusters en :
  - Cluster 1= pays consommant beaucoup de viande avec forte croissance démographique
  - Cluster 2=Pays consommant moyennement de viande avec une bonne croissance démographique
  - Cluster 3=pays consommant beaucoup de viande avec un faible taux de croissance démographique
  - cluster 4=Pays consommant moyennement de la viande avec un bon taux de croissance démographique et potentiellement riche pays
  - cluster 5= pays consommant beaucoup de viande avec un taux faible de croissance démographique



# RESULTATS

## CLASSEMENT

	CLUSTER	CRITERES
+++	1	-Consomme beaucoup de viande -Forte croissance démographique
+ - +	2	-Consomme moyennement de viande -Bonne croissance démographique
+ - -	3	-Consomme beaucoup de viande -Faible croissance démographique
+ - +	4	-Consomme moyennement de viande -Bonne croissance démographique
+ - -	5	-Consomme beaucoup de viande -Faible croissance démographique

# RESULTATS

## CHOIX DU CLUSTER I

---

	Zone	ratio prot	ratio ocr	disponibilite par habitant prot(kg/an/hab)	Disponibilité alimentaire (Kcal/personne/jour)	cluster
89	Maldives	60.443350	49.565217	29.63800	2235.0	1
5	Angola	30.744581	43.489195	19.36325	2383.0	1
81	Liban	32.541636	42.243882	25.20325	2851.0	1
73	Jordanie	33.540642	36.994776	25.23610	2729.0	1
108	Oman	49.337243	32.957048	31.11625	2938.0	1

# RESULTATS

## CRITÈRES DE CLASSEMENT

---

- Nous avons discriminé les variables et retenu la variable de croissance démographique comme la plus pertinente
- Le cluster I est le groupe qui connaît un fort taux de croissance démographique
- Il regorge aussi d'un taux moyen de consommation de protéine animale
- Ainsi , nous proposons une liste de ces 5 pays avec les Maldives en tête

# INTERPRETATION

## TESTS STATISTIQUES/LOI NORMALE

---

- Question: les clusters obtenus sont-ils distincts ?
- Q1: Vérifions si une des 4 variables suit une distribution de type normale
- Hypothèse de travail: une variable suit la forme de distribution loi normale
- $H_0$ : la variable ratiocr suit la loi Normale
- $H_1$ : Aucune des 4 variables n'est distribuée conformément à la loi normale
- Fixer  $\alpha$ : seuil de signification à 0.05 et rejet de  $H_0 > \alpha$
- Test appliqué = Kolmogorov Smirnov, variable quantitative continue
- $P\text{-value} = 0.02 < 0.05$  donc nous ne rejetons pas  $H_0$

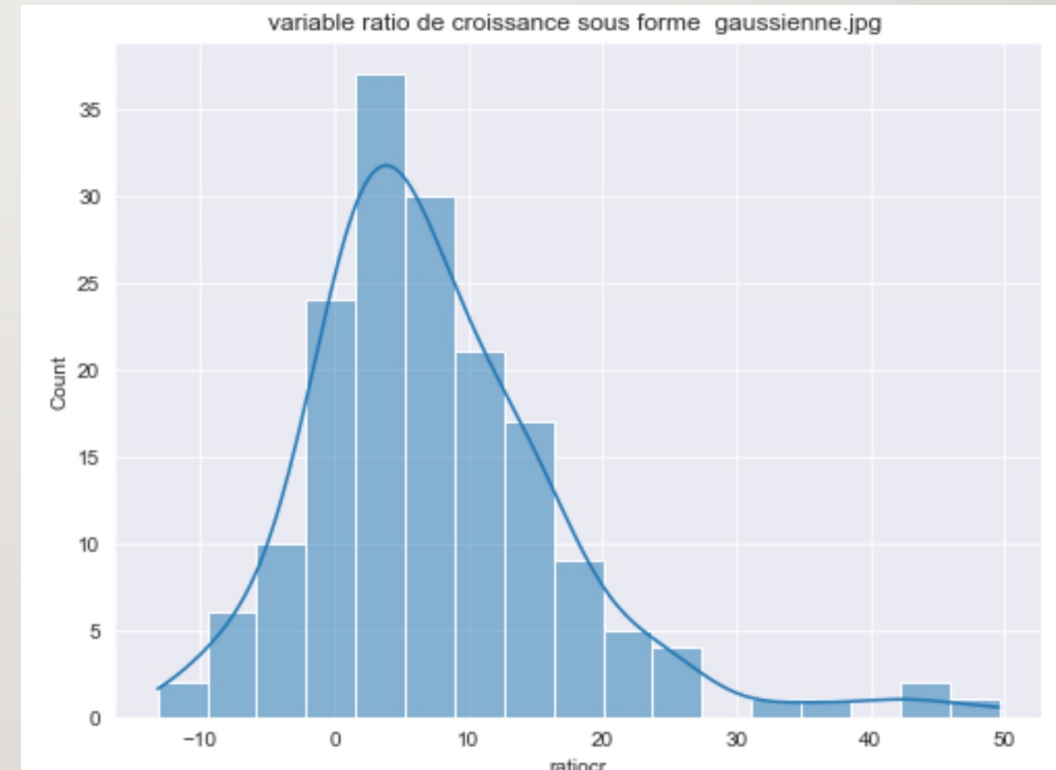


# TESTS STATISTIQUES

## ADÉQUATION LOI NORMALE

---

- La Var (ratiocr) est distribuée selon la loi normale
- Forme sous forme de cloche





- Nous allons tester la variable ratio de croissance en comparant deux clusters de notre échantillon .
- Nous décidons de tester les clusters 2 et 3
- Pour comparer deux échantillons dans notre cas les clusters à priori de loi gaussienne , il s'agira de tester l'égalité de leur variances et de leur moyennes

## TESTING TEST DE COMPARAISON

# TEST

## TEST DE COMPARAISON

---

- Nous allons tester l'égalité des variances et des moyennes
- Nous choisissons les clusters 2 et 3 et créons des variables pour calculer leurs variances et ensuite les comparer

$$\left\{ \begin{array}{l} H_0: \sigma^2_1 = \sigma^2_2 \quad H_1: \sigma^2_1 \neq \sigma^2_2 \end{array} \right.$$

Nous appliquons le test statistique avec la méthode bartlett et obtenons  $p\text{-value} = 0.1 > 0.05$

on ne rejette pas  $H_0$  car  $p\text{-value}$  est plus grand que notre seuil de test.

# INTERPRÉTATION TEST COMPARAISON

---

- Nous allons tester l'égalité des moyennes sachant que les variances sont égales
- $\begin{cases} H_0: \mu_1 = \mu_2 & H_1: \mu_1 \neq \mu_2 \end{cases}$
- Ce test d'égalité des moyennes indique une p-value faible que le seuil étudié., donc les moyennes dans les groupes sont différents. si nous posons l'hypothèse  $H_0: \text{moy}_1 = \text{moy}_2$  si p-value  $< 0.05$  alors que dans notre le p-value est très petit que 0.05 alors On peut donc très facilement rejeter à un niveau de test de 5%.

CONCLSION

# RECOMMANDATION

---

- Si l'étude de marché semble indiquer que le cluster I est porteur pour l'exportation de notre poulet ,il convient de prendre en compte d'autres paramètres au sein des clusters:
  - Les dérivés possibles du poulet et de l'aviculture (œufs ,nuggets ,filiales...)
  - L'aviculture dans les pays ciblés –commercialisations ,législation ...
  - les coûts liés à La distance géographique entre les pays d'importation et exportation
  - Ajouter à notre étude de marché une **analyse** marketing de type PESTEL