Projet d'Analyse des Données Qualitatives

Analyse Factorielle et Classification du Dataset Mushroom

[Nom Personne A] [Nom Personne B] [Nom Personne C]

Janvier 2026

Université [Nom]
Master [Spécialité]
Enseignant : [Nom Prof]

Abstract

Ce rapport présente une analyse complète du dataset UCI Mushroom, composé de 8 124 observations de champignons décrits par 23 variables qualitatives. L'objectif est de comprendre la structure des profils de champignons et d'identifier les caractéristiques discriminantes entre champignons comestibles et vénéneux. Nous appliquons une démarche structurée : statistiques descriptives, Analyse des Correspondances Multiples (ACM), classification non supervisée sur les composantes factorielles, et analyse discriminante. Les résultats révèlent que les caractéristiques de surface, l'odeur, et la couleur constituent les principaux axes de variation, permettant une discrimination efficace entre les deux classes.

Mots-clés : Analyse des Correspondances Multiples, Classification non supervisée, Analyse discriminante, Données qualitatives, Champignons

Contents

1 Introduction **4**
1.1 Motivation de l'étude 4
1.2 Démarche analytique 4
1.3 Structure du rapport 5

2 Problème, objectif et démarche **5**
2.1 Problématique 5
2.2 Objectifs de l'étude 5
2.2.1 Objectif 1 : Exploration et réduction de dimensionnalité 5
2.2.2 Objectif 2 : Segmentation non supervisée 6
2.2.3 Objectif 3 : Modélisation supervisée 6
2.3 Démarche globale : pipeline analytique 6

3 Données : description et préparation **7**
3.1 Source et contenu du dataset 7
3.1.1 Origine 7
3.1.2 Dimensions 7
3.1.3 Variable cible 7
3.2 Dictionnaire des variables 8
3.3 Nettoyage et préparation des données 9
3.3.1 Gestion des valeurs manquantes 9
3.3.2 Vérification de la qualité 9
3.4 Statistiques descriptives 9
3.4.1 Analyse univariée : distributions des variables clés 9
3.4.2 Analyse bivariée : relations avec la classe 10
3.5 Synthèse 11

4 Analyse des Correspondances Multiples (ACM) **11**
4.1 Rappel méthodologique 11
4.1.1 Principe de l'ACM 11
4.1.2 Interprétation des résultats 12
4.2 Résultats globaux : choix du nombre d'axes 12
4.2.1 Tableau des valeurs propres 12
4.2.2 Scree plot et règle de Kaiser 13

4.3 Interprétation des axes factoriels 13
4.3.1 Axe 1 (7,59%) : « Caractéristiques de surface et anneau » 13
4.3.2 Axe 2 (6,91%) : « Modalités rares et attachement des lamelles » . . 14
4.3.3 Axes 3 à 5 : compléments d'information 15
4.4 Projections et visualisations 15
4.4.1 Plan factoriel des modalités (axes 1-2) 15
4.4.2 Plan factoriel des individus (axes 1-2) 16
4.5 Export des coordonnées et bilan 17

5 Classification non supervisée sur composantes ACM 18

6 Analyse discriminante sur composantes ACM 18

7 Conclusion 18

A Dictionnaire complet des variables 20

B Tableaux de contributions et cos² complets 21

C Détails des résultats de clustering 21

D Détails des résultats de l'analyse discriminante 21

1 Introduction

L'identification des champignons comestibles constitue un enjeu majeur en mycologie, où une erreur de classification peut avoir des conséquences graves. Le dataset UCI Mushroom, issu du *Audubon Society Field Guide to North American Mushrooms* (1981), regroupe 8 124 observations de champignons décrits par 23 variables qualitatives morphologiques, telles que la forme du chapeau, l'odeur, la couleur des lamelles, ou encore la surface du pied.

1.1 Motivation de l'étude

Ce jeu de données présente plusieurs caractéristiques qui en font un cas d'étude idéal pour l'analyse de données qualitatives :

- **Richesse des variables** : 23 attributs qualitatifs couvrant différents aspects morphologiques
- **Taille substantielle** : Plus de 8 000 individus permettant des analyses statistiques robustes
- **Problématique binaire claire** : Classification edible (comestible) vs. poisonous (vénéneux)
- **Complexité des relations** : Interactions multiples entre variables morphologiques

1.2 Démarche analytique

Conformément aux exigences du projet, nous appliquons une méthodologie structurée en quatre étapes :

1. **Statistiques descriptives** : Exploration univariée et bivariée pour comprendre la distribution des variables et identifier les modalités dominantes
2. **Analyse des Correspondances Multiples (ACM)** : Réduction de dimensionnalité permettant de visualiser les profils de champignons dans un espace factoriel et d'identifier les axes de variation principaux
3. **Classification non supervisée** : Clustering (CAH et K-means) sur les coordonnées factorielles pour identifier des groupes naturels de champignons partageant des profils similaires
4. **Analyse discriminante** : Modélisation supervisée sur les composantes ACM pour expliquer et prédire la classe edible/poisonous, dans l'esprit de l'approche DISQUAL

Cette démarche permet de combiner exploration (non supervisée) et prédiction (supervisée) tout en valorisant la richesse des données qualitatives.

1.3 Structure du rapport

Le rapport s'organise comme suit : la Section 2 définit le problème, les objectifs et la démarche complète ; la Section 3 présente les données, le dictionnaire des variables et les statistiques descriptives ; la Section 4 détaille l'ACM et l'interprétation des axes factoriels ; les Sections 5 et 6 (réalisées par Personne B) couvrent respectivement le clustering et l'analyse discriminante ; enfin, la Section 7 conclut avec une synthèse des résultats et des perspectives.

2 Problème, objectif et démarche

2.1 Problématique

La classification des champignons repose sur l'expertise de mycologues capables d'identifier les espèces à partir de caractéristiques morphologiques. Cependant, cette expertise n'est pas universellement accessible, et les erreurs d'identification peuvent être fatales. La question centrale de cette étude est :

> *Peut-on identifier des profils-types de champignons et déterminer quelles caractéristiques morphologiques sont les plus discriminantes pour distinguer champignons comestibles et vénéneux ?*

Cette problématique se décline en trois sous-questions :

- **Structure latente** : Existe-t-il des axes de variation principaux qui résument l'information contenue dans les 23 variables ?
- **Groupes naturels** : Les champignons se regroupent-ils naturellement en clusters homogènes ?
- **Discrimination** : Quelles variables/modalités permettent de prédire efficacement la comestibilité ?

2.2 Objectifs de l'étude

2.2.1 Objectif 1 : Exploration et réduction de dimensionnalité

L'ACM vise à :

- Réduire la dimensionnalité de 23 variables à quelques axes factoriels interprétables
- Identifier les oppositions majeures entre modalités (e.g., odeur forte vs. absence d'odeur)
- Visualiser les profils de champignons dans un espace factoriel

2.2.2 Objectif 2 : Segmentation non supervisée

Le clustering sur composantes ACM permet de :

- Découvrir des groupes homogènes sans utiliser l'information de classe
- Profiler chaque cluster (modalités sur/sous-représentées)
- Évaluer la concordance entre clusters et classes edible/poisonous

2.2.3 Objectif 3 : Modélisation supervisée

L'analyse discriminante sur facteurs ACM (approche DISQUAL) vise à :

- Construire un modèle prédictif de la comestibilité basé sur les composantes
- Quantifier la performance de classification (matrice de confusion, taux de succès)
- Identifier les axes factoriels les plus discriminants

2.3 Démarche globale : pipeline analytique

La Figure 1 illustre la chaîne de traitement complète, de la collecte des données aux résultats finaux.

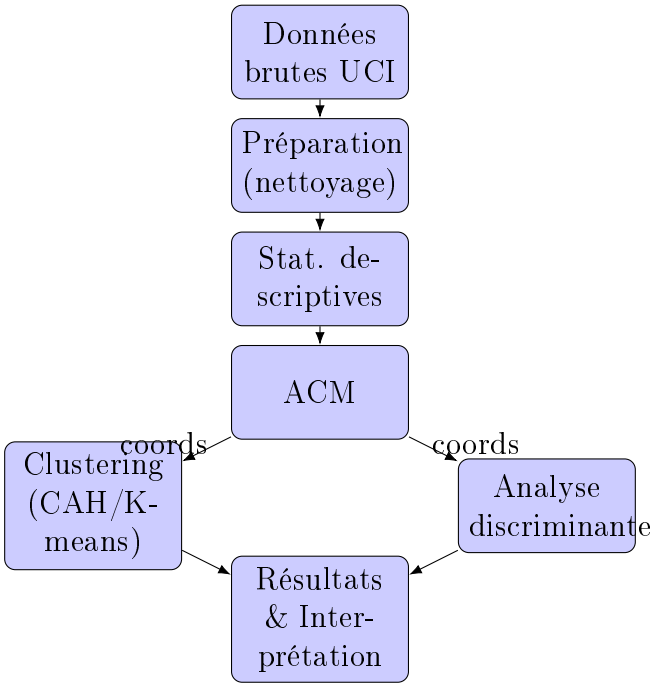


Figure 1: Pipeline analytique : de la collecte des données à l'interprétation finale

Justification de la démarche : L'utilisation de l'ACM comme étape intermédiaire est motivée par :

- **Nature des données** : Variables exclusivement qualitatives (incompatibles avec PCA)
- **Dimensionnalité** : 23 variables génèrent un tableau disjonctif complet de grande dimension
- **Interprétabilité** : Les axes factoriels sont plus interprétables que les variables brutes
- **Efficacité** : Réduire le bruit et conserver l'information discriminante

3 Données : description et préparation

3.1 Source et contenu du dataset

3.1.1 Origine

Le dataset *UCI Mushroom* provient du *UC Irvine Machine Learning Repository*. Il a été construit à partir du guide *Audubon Society Field Guide to North American Mushrooms* (1981) et décrit 23 espèces de champignons à lamelles des familles *Agaricus* et *Lepiota*.

3.1.2 Dimensions

- **Nombre d'individus** : $n = 8\ 124$ champignons
- **Nombre de variables** : $p = 23$ variables qualitatives (dont 1 variable cible)
- **Conformité** : Le dataset respecte largement les contraintes du projet ($n \geq 150$ et $p \geq 10$)

3.1.3 Variable cible

`class` : Indique si le champignon est comestible (`e`) ou vénéneux (`p`).

Table 1: Distribution de la variable cible

Classe	**Effectif**	**Pourcentage**
Comestible (e)	4 208	51,8%
Vénéneux (p)	3 916	48,2%
Total	**8 124**	**100,0%**

Le dataset est quasi-équilibré, ce qui élimine les problèmes de classes déséquilibrées pour la modélisation supervisée.

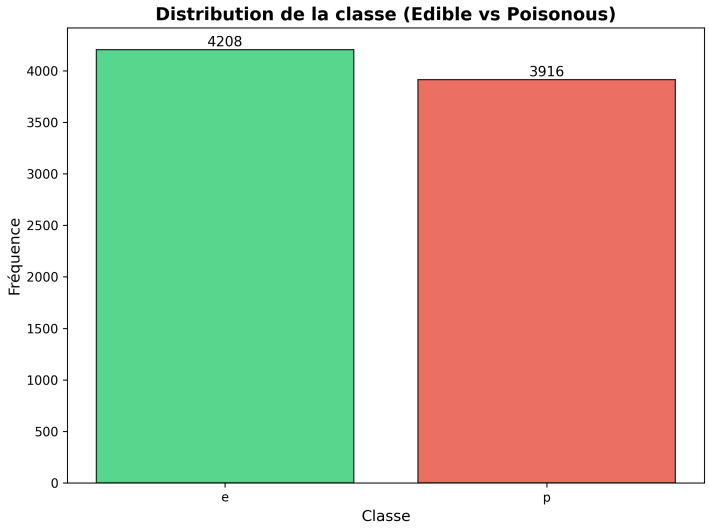


Figure 2: Distribution de la classe (Comestible vs. Vénéneux)

3.2 Dictionnaire des variables

Le Tableau 2 présente un résumé des 23 variables morphologiques. Le dictionnaire complet avec toutes les modalités est fourni en Annexe A.

Table 2: Résumé des variables du dataset (Top 10)

Variable	N_{mod}	**Modalité top**	**Freq. (%)**	**NA (%)**
class	2	e	51,8	0,0
cap-shape	6	x (convex)	45,0	0,0
cap-surface	4	y (scaly)	39,9	0,0
cap-color	10	n (brown)	28,1	0,0
bruises	2	f (no)	58,4	0,0
odor	9	n (none)	43,4	0,0
gill-attachment	2	f (free)	97,4	0,0
gill-spacing	2	c (close)	83,9	0,0
gill-size	2	b (broad)	69,1	0,0
gill-color	12	b (buff)	21,3	0,0

Observations clés :

- Certaines variables sont très déséquilibrées (`gill-attachment` : 97,4% de modalité `f`)
- Le nombre de modalités varie de 2 à 12 selon les variables
- La variable `stalk-root` contient 30,5% de valeurs manquantes (voir section suivante)

3.3 Nettoyage et préparation des données

3.3.1 Gestion des valeurs manquantes

La variable `stalk-root` présente 2 480 valeurs manquantes (30,5% du dataset), encodées par le symbole « ? » dans les données brutes.

Stratégie retenue : Remplacement par imputation modale

- Les valeurs « ? » sont remplacées par la modalité la plus fréquente de `stalk-root` (`b` : bulbous)
- **Justification** : Cette approche préserve la distribution majoritaire tout en permettant l'utilisation de tous les individus dans l'ACM
- **Alternative** : Suppression des lignes (perte de 30% des données) ou création d'une modalité « missing » (augmente artificiellement le nombre de modalités)

3.3.2 Vérification de la qualité

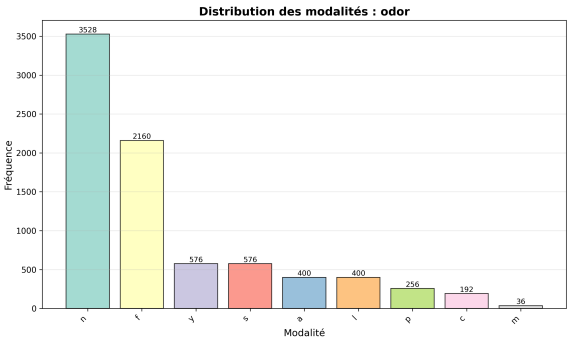
Après nettoyage :

- **Aucune valeur manquante résiduelle**
- **Toutes les variables sont de type qualitatif** (nominal/ordinal)
- **Pas de modalité ultra-rare** nécessitant un regroupement (seuil : >0,5%)

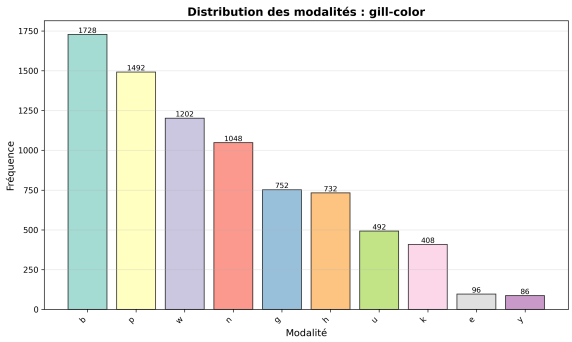
3.4 Statistiques descriptives

3.4.1 Analyse univariée : distributions des variables clés

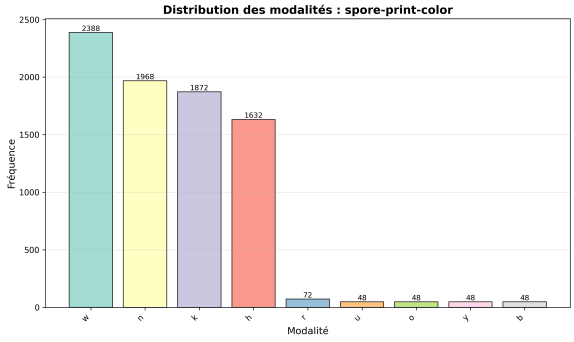
Nous présentons ici la distribution de trois variables morphologiques d'intérêt : `odor`, `gill-color`, et `spore-print-color`.



(a) Odeur



(b) Couleur des lamelles



(c) Couleur de l'empreinte des spores

Figure 3: Distributions des modalités pour trois variables morphologiques clés

Observations :

- **Odeur** : Près de 43% des champignons n'ont pas d'odeur (`n`), tandis que les odeurs fortes (`f` : foul, `p` : pungent) concernent environ 25% des individus
- **Couleur des lamelles** : Distribution très fragmentée avec 12 modalités, la plus fréquente étant `b` (buff) à 21%
- **Empreinte des spores** : Dominée par le blanc (`w`, 29%), suivie par le marron (`n`, 24%) et le noir (`k`, 19%)

3.4.2 Analyse bivariée : relations avec la classe

Le Tableau 3 présente le tableau croisé entre `odor` et `class`, révélant une association forte.

Table 3: Tableau croisé : Odeur × Classe

Odeur	Comestible	Vénéneux	Total
n (none)	3 408	120	3 528
f (foul)	0	2 160	2 160
a (almond)	400	0	400
l (anise)	400	0	400
p (pungent)	0	256	256
Autres	0	1 380	1 380
Total	**4 208**	**3 916**	**8 124**

Analyse : Les odeurs `a` (almond) et `l` (anise) sont exclusivement associées aux champignons comestibles, tandis que `f` (foul) et `p` (pungent) sont des indicateurs quasi-parfaits de toxicité. Cette variable sera donc probablement très discriminante dans l'analyse factorielle et la modélisation supervisée.

3.5 Synthèse

Le dataset Mushroom est de haute qualité :

- **Taille robuste** : 8 124 observations permettent des analyses statistiques fiables
- **Richesse** : 23 variables qualitatives couvrant différents aspects morphologiques
- **Équilibre** : Classes quasi-équilibrées (51,8% vs. 48,2%)
- **Propreté** : Après nettoyage, aucune valeur manquante résiduelle

Les statistiques descriptives révèlent déjà des patterns prometteurs : certaines variables (`odor`, `spore-print-color`) semblent fortement associées à la comestibilité. L'ACM (Section 4) permettra de synthétiser ces informations et de révéler la structure latente des données.

4 Analyse des Correspondances Multiples (ACM)

4.1 Rappel méthodologique

4.1.1 Principe de l'ACM

L'Analyse des Correspondances Multiples est une technique de réduction de dimensionnalité adaptée aux variables qualitatives. Elle généralise l'Analyse Factorielle des Correspondances (AFC) au cas de p variables qualitatives.

Construction du tableau disjonctif complet (TDC) :

- Chaque variable V_j à k_j modalités est transformée en k_j variables indicatrices binaires

- Le TDC résultant a n lignes (individus) et $K = \sum_{j=1}^{p} k_j$ colonnes (modalités)
- Dans notre cas : $p = 22$ variables (hors `class`) $\rightarrow K = 111$ modalités au total

Diagonalisation : L'ACM effectue une diagonalisation du tableau de Burt (matrice $K \times K$ des croisements de modalités) et extrait les axes factoriels maximisant l'inertie expliquée.

Inertie totale : Dans une ACM, l'inertie totale vaut :

$$I_{tot} = \frac{p - K/p}{p} = \frac{22 - 111/22}{22} \approx 4.27$$

4.1.2 Interprétation des résultats

- **Valeurs propres** : Mesurent la part de variance expliquée par chaque axe
- **Contributions** : Identifient les modalités qui « pèsent » le plus sur un axe
- **Cos²** (qualité de représentation) : Mesure la qualité de projection d'une modalité/individu sur un axe

4.2 Résultats globaux : choix du nombre d'axes

4.2.1 Tableau des valeurs propres

Le Tableau 4 présente les 10 premières valeurs propres et les inerties expliquées associées.

Table 4: Valeurs propres et inerties expliquées (ACM)

Axe	Valeur propre	Inertie (%)	Inertie cum. (%)
Dim1	0,324	7,59	7,59
Dim2	0,295	6,91	14,49
Dim3	0,271	6,33	20,83
Dim4	0,243	5,68	26,51
Dim5	0,203	4,76	**31,27**
Dim6	0,193	4,51	35,78
Dim7	0,173	4,06	39,83
Dim8	0,144	3,38	43,21
Dim9	0,103	2,42	45,62
Dim10	0,096	2,25	47,88

Remarque : Les inerties individuelles sont faibles ($< 10\%$), ce qui est typique en ACM lorsque le nombre de modalités K est élevé. L'inertie se disperse sur de nombreux axes.

4.2.2 Scree plot et règle de Kaiser

La Figure 4 présente le scree plot et l'inertie cumulée.

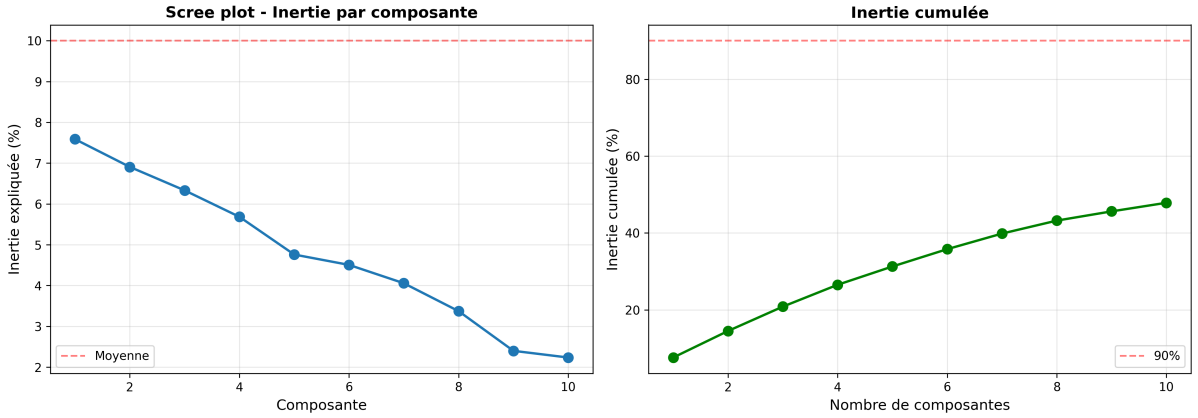


Figure 4: Scree plot : inertie par composante (gauche) et inertie cumulée (droite)

Analyse :

- **Coude** : Un coude est visible après les axes 5-6, suggérant que les axes suivants apportent peu d'information supplémentaire
- **Règle des 90%** : Il faudrait plus de 20 axes pour atteindre 90% d'inertie cumulée (non réaliste)
- **Compromis retenu** : Nous conservons $\mathbf{k = 5}$ **axes** (31,27% d'inertie cumulée), offrant un équilibre entre interprétabilité et conservation de l'information

4.3 Interprétation des axes factoriels

Nous interprétons ici les deux premiers axes, qui concentrent 14,49% de l'inertie totale.

4.3.1 Axe 1 (7,59%) : « Caractéristiques de surface et anneau »

Le Tableau 5 présente les modalités contribuant le plus à l'axe 1.

Table 5: Top 10 des contributions à l'axe 1

Modalité	Coordonnée	Contribution (%)
ring-type__l (large ring)	1,728	6,68
stalk-surface-below-ring__k (silky)	1,270	6,42
stalk-surface-above-ring__k (silky)	1,219	6,09
odor__f (foul)	1,213	5,49
spore-print-color__h (chocolate)	1,333	5,01
ring-type__p (pendant)	-0,667	3,05
bruises__t (bruises present)	-0,654	2,49
stalk-color-below-ring__b (buff)	1,796	2,41
odor__n (no odor)	-0,622	2,36

Interprétation :

Pôle positif (coordonnées > 0) : Champignons avec anneau large (`ring-type__l`), surface du pied soyeuse (`stalk-surface-*__k`), odeur forte et désagréable (`odor__f`), empreinte de spores chocolat (`spore-print-color__h`).

Pôle négatif (coordonnées < 0) : Champignons avec anneau pendant (`ring-type__p`), présence de bleus (`bruises__t`), absence d'odeur (`odor__n`).

Sens de l'axe : L'axe 1 oppose les champignons à **surface lisse/soyeuse** + **odeur forte** (pôle positif) aux champignons à **anneau pendant** + **bleus** (pôle négatif). Cet axe capture donc les caractéristiques de texture de surface et d'odeur.

Lien avec la classe : Les modalités du pôle positif (`odor__f`, etc.) sont majoritairement associées aux champignons vénéneux (cf. Tableau 3), suggérant que cet axe a un pouvoir discriminant.

4.3.2 Axe 2 (6,91%) : « Modalités rares et attachement des lamelles »

Le Tableau 6 présente les modalités contribuant le plus à l'axe 2.

Table 6: Top 10 des contributions à l'axe 2

Modalité	Coordonnée	Contribution (%)
gill-attachment__a (attached)	4,675	8,70
stalk-color-below-ring__o (orange)	4,434	7,16
stalk-color-above-ring__o (orange)	4,434	7,16
habitat__l (leaves)	1,692	4,52
population__c (clustered)	2,625	4,44
gill-color__y (yellow)	5,175	4,37
veil-color__n (brown)	4,435	3,58
veil-color__o (orange)	4,434	3,58

Interprétation :

L'axe 2 est dominé par des **modalités rares** (e.g., `gill-attachment__a` : 3% des individus, `stalk-color-*__o` : <1%). Ces modalités ont des coordonnées très élevées car

elles sont éloignées du centre de gravité.

Sens de l'axe : L'axe 2 oppose les champignons présentant des caractéristiques atypiques (lamelles attachées, couleurs orange/jaune, habitat spécifique) aux champignons « moyens » (modalités fréquentes).

Limite : Cet axe reflète principalement des **effets de taille** (modalités rares vs. fréquentes) plutôt qu'une opposition sémantique forte. Cependant, il capture une partie de la variabilité intra-espèces et peut être utile pour identifier des sous-groupes spécifiques.

4.3.3 Axes 3 à 5 : compléments d'information

Les axes 3 à 5 (non détaillés ici) capturent des nuances supplémentaires :

- **Axe 3** : Variations de couleur du chapeau et de la tige
- **Axe 4** : Opposition entre habitats (bois vs. prairies)
- **Axe 5** : Forme du pied (élargissement vs. effilé)

Ces axes seront utilisés dans le clustering et l'analyse discriminante mais ne seront pas interprétés exhaustivement dans ce rapport.

4.4 Projections et visualisations

4.4.1 Plan factoriel des modalités (axes 1-2)

La Figure 5 projette les 111 modalités sur le plan factoriel 1-2.

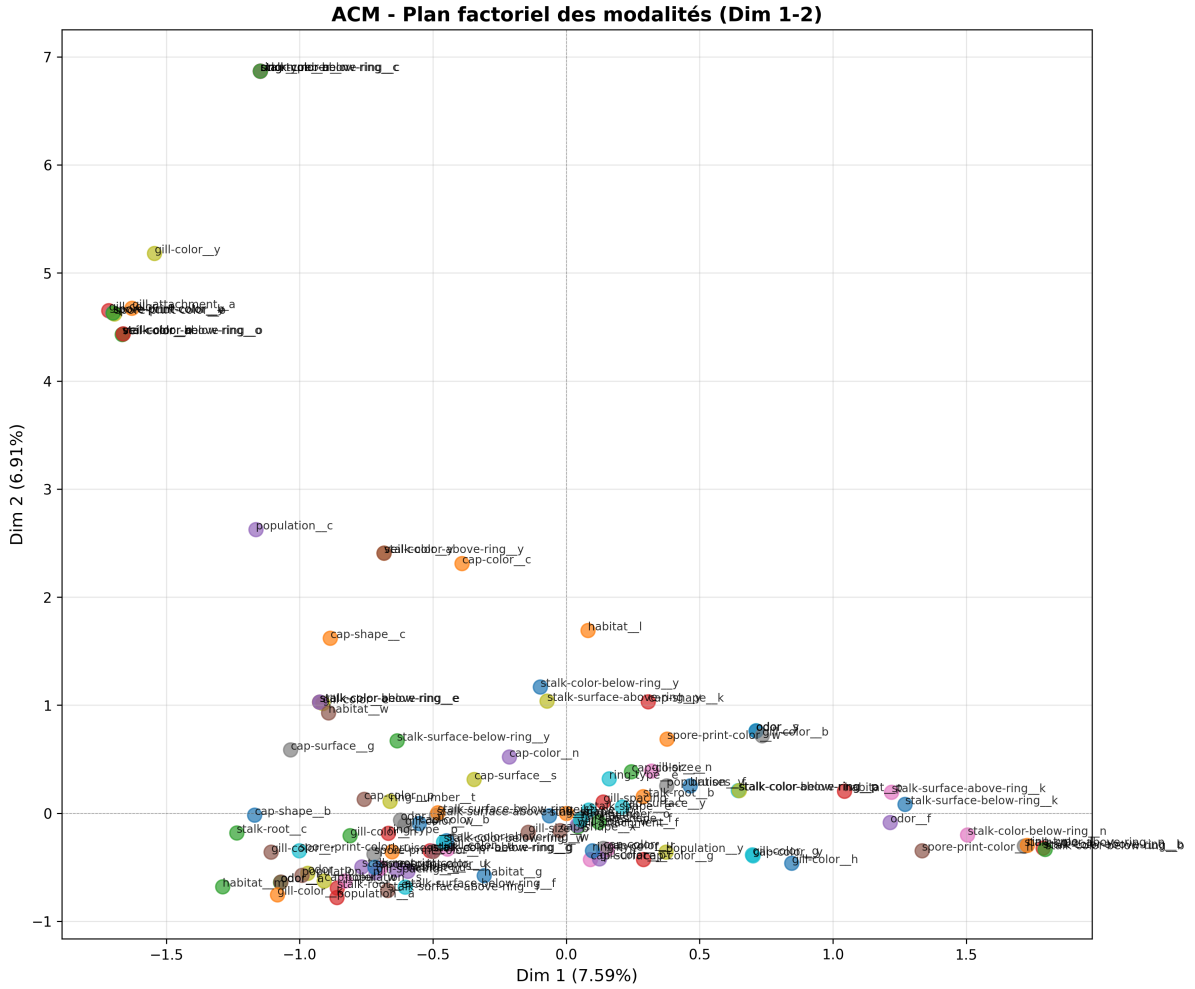


Figure 5: Plan factoriel des modalités (axes 1-2)

Observations :

- **Dispersion sur l'axe 1** : Opposition entre modalités à gauche (`odor__n`, `ring-type__p`) et à droite (`odor__f`, `ring-type__l`)
- **Modalités excentrées sur l'axe 2** : Les modalités rares (`gill-attachment__a`, `gill-color__y`) sont très éloignées du centre
- **Centre de gravité** : Les modalités fréquentes (e.g., `gill-attachment__f`, `cap-shape__x`) sont proches de l'origine

4.4.2 Plan factoriel des individus (axes 1-2)

La Figure 6 projette les 8 124 individus (champignons) sur le plan 1-2, colorés selon leur classe (comestible/vénéneux).

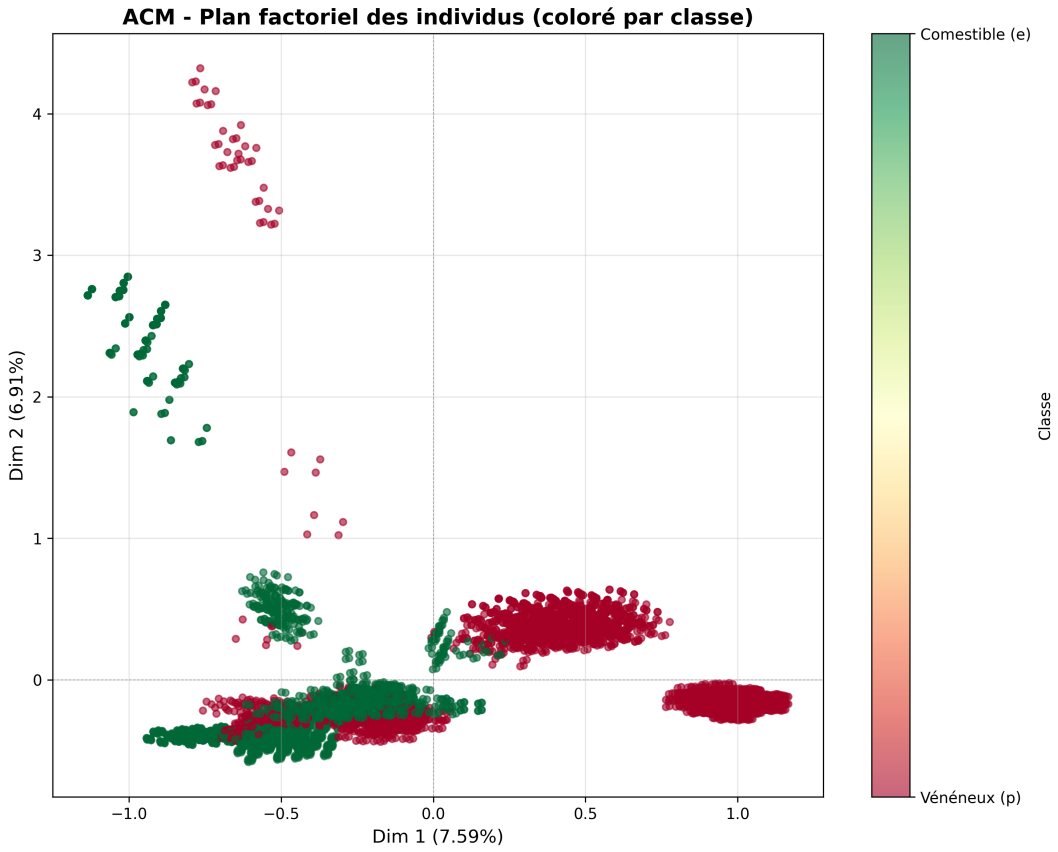


Figure 6: Plan factoriel des individus (axes 1-2), colorés par classe

Analyse :

- **Séparation partielle** : On observe une tendance à la séparation entre comestibles (vert) et vénéneux (rouge) le long de l'axe 1, cohérente avec l'interprétation (odeur, anneau)
- **Superposition** : Les nuages se chevauchent fortement, indiquant que les axes 1-2 seuls ne suffisent pas pour une discrimination parfaite
- **Axes 3-5** : L'utilisation des 5 premiers axes (31,27% d'inertie) devrait améliorer la séparation (à vérifier dans l'analyse discriminante, Section 6)

4.5 Export des coordonnées et bilan

Les coordonnées des 8 124 individus sur les 10 premiers axes factoriels ont été exportées dans le fichier `mca_coords.csv` (disponible avec le rapport). Ce fichier servira d'entrée pour :

- La classification non supervisée (Section 5) : clustering sur les k = 5 premières colonnes
- L'analyse discriminante (Section 6) : modélisation supervisée sur les mêmes 5 axes

Bilan de l'ACM :

- **Réduction de dimensionnalité** : De 22 variables (111 modalités) à 5 axes factoriels interprétables
- **Structure révélée** : Deux axes principaux capturant respectivement les caractéristiques de surface/odeur (axe 1) et les modalités rares (axe 2)
- **Pouvoir discriminant** : L'axe 1 montre une séparation partielle entre classes, prometteur pour la suite
- **Limite** : L'inertie expliquée reste modérée (31,27%), typique en ACM avec de nombreuses variables

Les Sections 5 et 6 (réalisées par Personne B) exploiteront ces coordonnées factorielles pour (i) découvrir des groupes naturels de champignons et (ii) construire un modèle prédictif de la comestibilité.

5 Classification non supervisée sur composantes ACM

[Cette section sera rédigée par Personne B]

6 Analyse discriminante sur composantes ACM

[Cette section sera rédigée par Personne B]

7 Conclusion

[À compléter après intégration des sections 5-6]

Cette étude a permis d'analyser de manière approfondie le dataset UCI Mushroom à travers une démarche structurée combinant exploration (ACM, clustering) et prédiction (analyse discriminante).

Principaux résultats :

- L'ACM a révélé deux axes majeurs : (1) caractéristiques de surface et odeur, (2) modalités rares

résultats clustering

ux de classification

Apports de l'analyse :

- Identification des variables morphologiques discriminantes pour la comestibilité

- Compréhension de la structure latente des profils de champignons

À compléter

Limites :

- L'inertie expliquée par l'ACM reste modérée (31% sur 5 axes)
- Certaines modalités très rares génèrent des effets de taille sur l'axe 2
- Le dataset ne couvre que des champignons à lamelles (non généralisable à toutes les espèces)

Perspectives :

- Tester d'autres méthodes de réduction (FAMD si variables mixtes, t-SNE pour visualisation)
- Comparer l'approche DISQUAL avec d'autres classifieurs (Random Forest, SVM)
- Étendre l'analyse à d'autres datasets mycologiques

References

[1] Dua, D. and Graff, C. (2019). *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science. http://archive.ics.uci.edu/ml

[2] Lincoff, G. H. (1981). *The Audubon Society Field Guide to North American Mushrooms*. Alfred A. Knopf, New York.

[3] Lebart, L., Morineau, A., et Piron, M. (2006). *Statistique exploratoire multidimensionnelle*. 4e édition, Dunod.

[4] Pagès, J. (2014). *Multiple Factor Analysis by Example Using R*. Chapman & Hall/CRC.

[5] Saporta, G. (2011). *Probabilités, analyse des données et statistique*. 3e édition, Éditions Technip.

A Dictionnaire complet des variables

Table 7: Dictionnaire exhaustif des 23 variables

Variable	Description et modalités
`class`	Classe : `e` = edible (comestible), `p` = poisonous (vénéneux)
`cap-shape`	Forme du chapeau : `b` = bell, `c` = conical, `x` = convex, `f` = flat, `k` = knobbed, `s` = sunken
`cap-surface`	Surface du chapeau : `f` = fibrous, `g` = grooves, `y` = scaly, `s` = smooth
`cap-color`	Couleur du chapeau : `n` = brown, `b` = buff, `c` = cinnamon, `g` = gray, `r` = green, `p` = pink, `u` = purple, `e` = red, `w` = white, `y` = yellow
`bruises`	Présence de bleus : `t` = bruises, `f` = no bruises
`odor`	Odeur : `a` = almond, `l` = anise, `c` = creosote, `y` = fishy, `f` = foul, `m` = musty, `n` = none, `p` = pungent, `s` = spicy
`gill-attachment`	Attachement des lamelles : `a` = attached, `d` = descending, `f` = free, `n` = notched
`gill-spacing`	Espacement des lamelles : `c` = close, `w` = crowded, `d` = distant
`gill-size`	Taille des lamelles : `b` = broad, `n` = narrow
`gill-color`	Couleur des lamelles : `k` = black, `n` = brown, `b` = buff, `h` = chocolate, `g` = gray, `r` = green, `o` = orange, `p` = pink, `u` = purple, `e` = red, `w` = white, `y` = yellow
`stalk-shape`	Forme du pied : `e` = enlarging, `t` = tapering
`stalk-root`	Racine du pied : `b` = bulbous, `c` = club, `u` = cup, `e` = equal, `z` = rhizomorphs, `r` = rooted, `?` = missing
`stalk-surface-above-ring`	Surface pied au-dessus anneau : `f` = fibrous, `y` = scaly, `k` = silky, `s` = smooth
`stalk-surface-below-ring`	Surface pied en-dessous anneau : `f` = fibrous, `y` = scaly, `k` = silky, `s` = smooth
`stalk-color-above-ring`	Couleur pied au-dessus anneau : `n` = brown, `b` = buff, `c` = cinnamon, `g` = gray, `o` = orange, `p` = pink, `e` = red, `w` = white, `y` = yellow
`stalk-color-below-ring`	Couleur pied en-dessous anneau : `n` = brown, `b` = buff, `c` = cinnamon, `g` = gray, `o` = orange, `p` = pink, `e` = red, `w` = white, `y` = yellow
`veil-type`	Type de voile : `p` = partial, `u` = universal
`veil-color`	Couleur du voile : `n` = brown, `o` = orange, `w` = white, `y` = yellow
`ring-number`	Nombre d'anneaux : `n` = none, `o` = one, `t` = two
`ring-type`	Type d'anneau : `c` = cobwebby, `e` = evanescent, `f` = flaring, `l` = large, `n` = none, `p` = pendant, `s` = sheathing, `z` = zone

(suite page suivante)

(suite)

Variable	Description et modalités
`spore-print-color`	Couleur empreinte spores : `k` = black, `n` = brown, `b` = buff, `h` = chocolate, `r` = green, `o` = orange, `u` = purple, `w` = white, `y` = yellow
`population`	Population : `a` = abundant, `c` = clustered, `n` = numerous, `s` = scattered, `v` = several, `y` = solitary
`habitat`	Habitat : `g` = grasses, `l` = leaves, `m` = meadows, `p` = paths, `u` = urban, `w` = waste, `d` = woods

B Tableaux de contributions et cos² complets

[Tableaux détaillés des contributions des modalités aux axes 1 à 5, et qualités de représentation]

C Détails des résultats de clustering

[Sera complété par Personne B : tableaux de profils complets, v-tests, etc.]

D Détails des résultats de l'analyse discriminante

[Sera complété par Personne B : coefficients, résultats de validation croisée détaillés, etc.]