Female Farmers' Health Study

Multivariate Analysis Report

Submitted by: Selma Bettaieb

Date: April 17, 2025

Table of Contents

Unveiling Hidden Patterns: Multivariate Analysis of Female Farmers' Health Study

Introduction and Rationale

Our exploration of the female farmers' health dataset has revealed numerous univariate and bivariate patterns, but the complexity of occupational health factors cannot be fully understood through these analyses alone. To uncover deeper, multidimensional relationships, we applied two complementary multivariate techniques:

1. **Principal Component Analysis (PCA)** - To identify latent dimensions in our numerical variables (age, work hours, health metrics)
2. **Multiple Correspondence Analysis (MCA)** - To discover hidden patterns in our categorical variables (education, protection equipment usage, exposure types)

This integrated approach provides several advantages over traditional analysis methods:

- **Dimensionality reduction**: Condenses dozens of variables into a few interpretable dimensions
- **Pattern recognition**: Reveals complex relationships invisible to univariate and bivariate analyses
- **Natural clustering**: Identifies distinct profiles of female farmers with similar characteristics
- **Integration of mixed data types**: Bridges numerical and categorical variables to provide holistic insights
- **Enhanced visualization**: Maps complex relationships into interpretable spatial representations

By combining these complementary techniques, we can move beyond simple correlations to understand the multidimensional nature of agricultural health risks and develop targeted interventions for specific farmer profiles.

Principal Component Analysis: Uncovering Numerical Dimensions

Why PCA for Our Dataset?

Our dataset contains 11 interrelated numerical variables (age, work hours, health metrics, etc.) that present several analytical challenges:

- **Correlation and redundancy** - Many variables measure related concepts (e.g., age correlates with years of experience at r=0.86)
- **Dimensionality issues** - Having numerous variables makes visualization and interpretation difficult

- **Noise in individual variables** - Individual measurements contain random variations that can obscure patterns
- **Multicollinearity** - Related variables (like weight and height) make it difficult to isolate independent effects
- **Interpretation complexity** - Understanding how 11 different variables interact simultaneously is cognitively challenging

PCA transforms our original variables into new composite variables (principal components) that:

- Capture maximum variance in the data in decreasing order of importance
- Are uncorrelated with each other (orthogonal)
- Represent underlying dimensions of variation in an interpretable way
- Simplify the data structure while preserving meaningful information

> Hint for Understanding PCA: Think of PCA like taking multiple photos of a 3D object from different angles. Each "photo" (principal component) captures a different perspective of the same data, with the first photo showing the angle with the most detail.

Methodology and Implementation

We applied PCA to the following standardized numerical variables:

- Demographic variables: Age, Nb enfants (number of children), Nb pers à charge (dependents)
- Work characteristics: H travail / jour (hours/day), J travail / Sem (days/week), Ancienneté agricole (years of experience)
- Health indicators: Poids (weight), Taille (height), TAS (systolic pressure), TAD (diastolic pressure), GAD (global arterial pressure)

The analysis followed these steps:

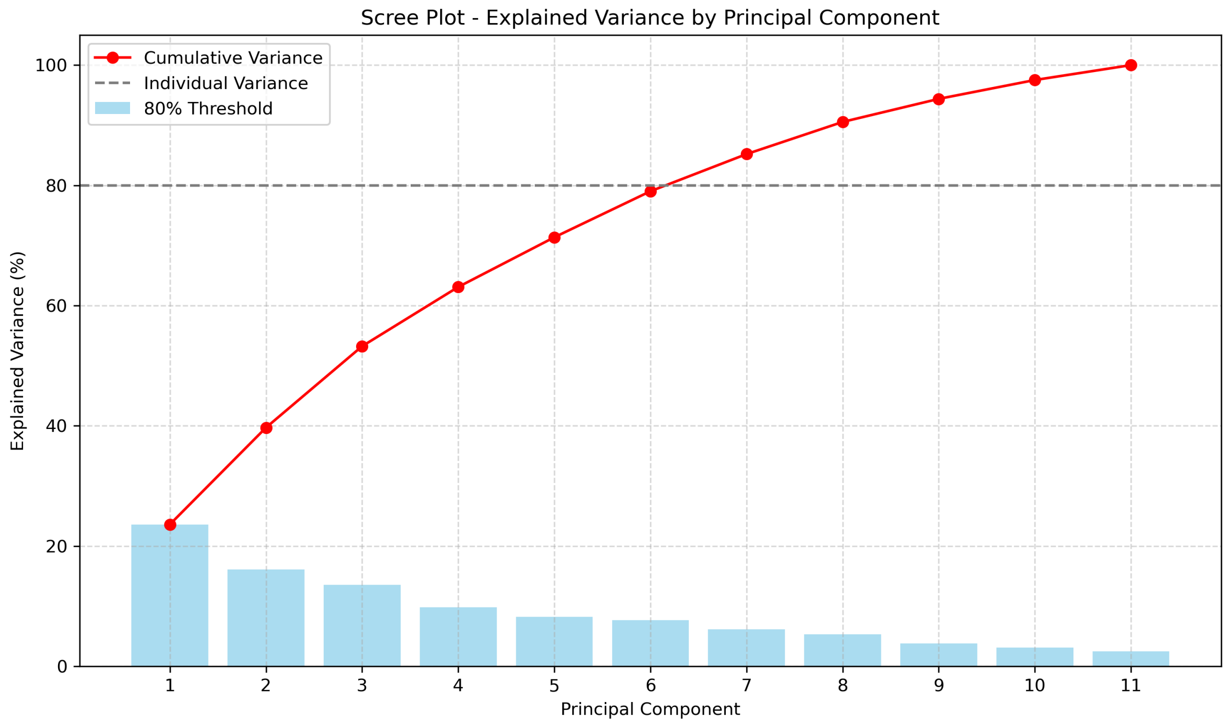
1. Data standardization to ensure comparability across different scales

2. Computation of the correlation matrix to identify relationships between variables
3. Extraction of eigenvalues and eigenvectors to determine the principal components
4. Interpretation of component loadings to understand what each dimension represents
5. Projection of observations onto the new dimensions for visualization and clustering

> Hint for Interpreting Correlation Heatmaps: The darker the blue, the stronger the positive correlation; the darker the red, the stronger the negative correlation. Look for clusters of correlated variables as they often reveal underlying dimensions.

Key PCA Results

Our analysis revealed four principal components that together explain 63.1% of the total variance:



> Hint for Reading Scree Plots: The "elbow" point where the curve flattens indicates a good cutoff for the number of components to retain. Components beyond this point typically add little additional information.

PC1: Age and Experience Factor (23.6% of variance)

- **Strong positive loadings**: Age (0.87), Ancienneté agricole (0.77), TAS (0.67)
- **Interpretation**: Represents the cumulative effects of aging and agricultural experience on cardiovascular health
- **Key insight**: This dimension captures the natural aging process in agricultural workers, where systolic blood pressure increases with age and experience

PC2: Family Structure Factor (16.1% of variance)

- **Strong negative loadings**: Nb enfants (-0.70), Nb pers à charge (-0.60), Taille (-0.59), Poids (-0.58)
- **Interpretation**: Represents an inverse relationship between family size/responsibilities and physical stature

- **Key insight**: Women with more children and dependents tend to have different physical characteristics, suggesting potential effects of reproductive history on physical attributes

PC3: Cardiovascular Health Factor (13.5% of variance)

- **Positive loadings**: GAD (0.42)
- **Negative loadings**: TAD (-0.75), Poids (-0.54)
- **Interpretation**: Contrasts global arterial pressure with diastolic pressure and weight
- **Key insight**: Reveals a complex physiological relationship where GAD increases while TAD and weight decrease, potentially indicating different cardiovascular risk profiles

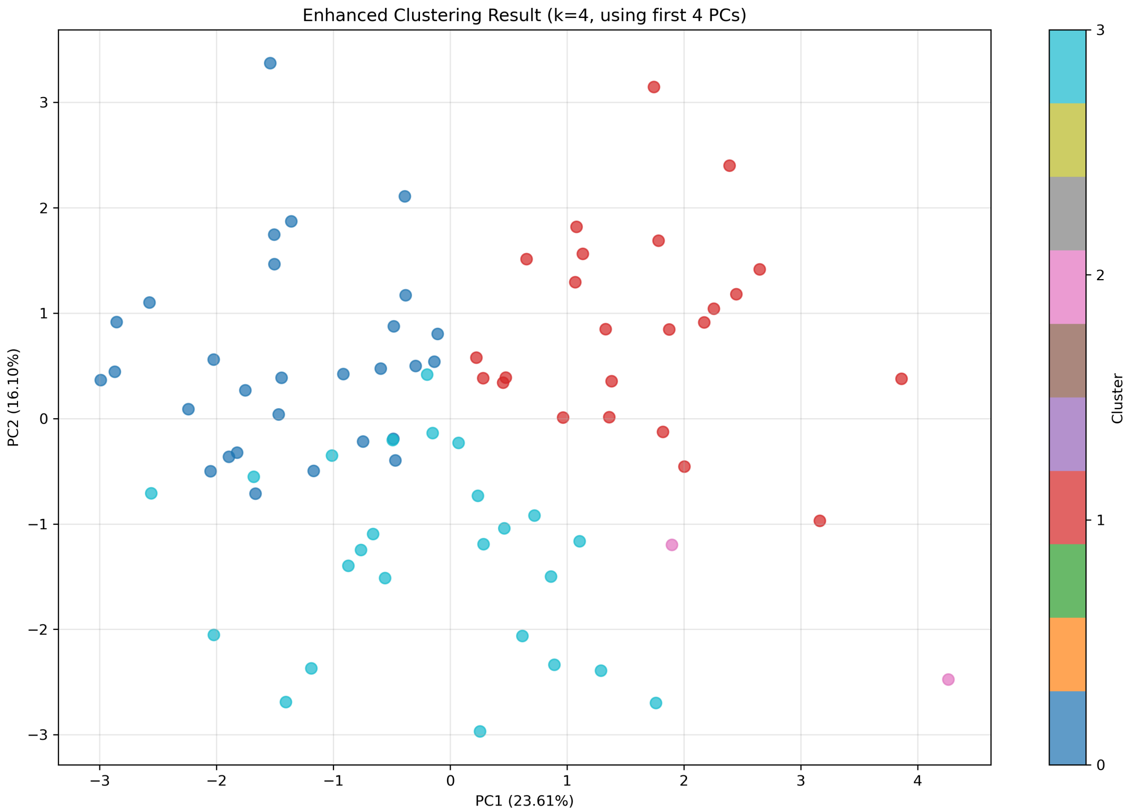
PC4: Work Intensity Factor (9.9% of variance)

- **Positive loadings**: TAS (0.49)
- **Negative loadings**: J travail / Sem (-0.52), H travail / jour (-0.45)
- **Interpretation**: Contrasts systolic blood pressure with work intensity (hours and days)
- **Key insight**: Unexpectedly shows an inverse relationship between work intensity and blood pressure, suggesting potential physiological adaptation to demanding work schedules

> Hint for Understanding Component Loadings: A high positive loading means the variable increases as the component increases. A high negative loading means the variable decreases as the component increases. Look for patterns of related variables loading onto the same component.

Enhanced Clustering Based on PCA Dimensions

To identify meaningful subgroups within our population, we applied K-means clustering to the principal components. After evaluating silhouette scores across different numbers of clusters, we determined that four clusters provided the optimal balance between differentiation and interpretability.



> Hint for Interpreting Cluster Plots: Clusters that are well-separated in this plot represent genuinely different farmer profiles. The axes (PC1 and PC2) provide the context for understanding what makes each cluster distinct - in this case, age/experience (PC1) and family structure (PC2).

The clustering revealed four distinct profiles of female agricultural workers, each with unique combinations of age, experience, family structure, and health characteristics:

Cluster 0: Younger Workers with Better Protection (36.2% of sample, n=29)

Demographic Profile:

- Significantly younger (38.5 years, -18.4% vs overall average)

- Substantially less agricultural experience (8.6 years, -50.7% vs overall)
- Fewer children (2.1 vs overall average of 2.4) and significantly fewer dependents (0.9, -36.9% vs overall)

Educational & Socioeconomic Status:

- Highest education levels among all clusters (58.6% secondary education, +13.6% vs overall)
- 17.2% have higher education (+6.0% vs overall)
- Predominantly middle socioeconomic status (75.9%, +10.9% vs overall)

Work Patterns:

- Less intensive work schedule (5.3 hours/day, -14.2%; 5.8 days/week, -10.0%)
- Higher proportion of seasonal workers (20.7%, +5.7% vs overall)

Health & Protection:

- Lower blood pressure indicators (TAS 113.1, -9.2%; TAD 70.7, -4.1%)
- Significantly higher protection equipment usage scores (+8.5% vs overall)
- Lower exposure to traditional practices (55.2% Tabouna exposure, -18.6% vs overall)

Interpretation: This cluster represents the youngest, best-educated segment of female agricultural workers who show the healthiest work-life balance and best protective behaviors despite less experience. They likely benefit from better awareness of occupational hazards due to their education level.

Cluster 1: Older High-Intensity Workers (30.0% of sample, n=24)

Demographic Profile:

- Significantly older (59.3 years, +25.8% vs overall)
- Extensive agricultural experience (27.8 years, +58.8% vs overall)
- Fewer children (1.6, -31.6% vs overall) but average number of dependents

Educational & Socioeconomic Status:

- Mixed education levels (37.5% secondary, -7.5%; 33.3% primary, +0.8%)

- Higher illiteracy rate (29.2%, +17.9% vs overall)
- More diverse marital status (58.3% married, -20.4%; 25.0% single, +11.2%)

Work Patterns:

- Highest work intensity of all clusters (7.3 hours/day, +18.2%; 6.9 days/week, +6.2%)
- Predominantly permanent workers (87.5%, +2.5% vs overall)

Health & Protection:

- Concerning cardiovascular indicators (TAS 138.3, +11.0%; TAD 79.2, +7.3%)
- Slightly lower protection scores (-2.2% vs overall)
- Very high exposure to traditional practices (87.5% Tabouna exposure, +13.8%)

Interpretation: This cluster represents veteran agricultural workers with the highest work demands and poorest work-life balance. Their high work intensity combined with elevated cardiovascular risk indicators and minimal protective behaviors suggests a vulnerable population that prioritizes productivity over safety.

Cluster 2: Elderly High-Risk Workers (2.5% of sample, n=2)

Demographic Profile:

- Eldest age group (67.0 years, +42.2% vs overall)
- Most extensive agricultural experience (42.5 years, +143.2%)
- Highest number of children (6.0, +152.6% vs overall)

Educational & Socioeconomic Status:

- Universal illiteracy (100% analphabète)
- Lower socioeconomic status (50% "bas", +31.2% vs overall)
- Higher widowhood rate (50%, +45.0% vs overall)

Work Patterns:

- Average work hours but maintaining full permanent status (100%)
- Continuing to work despite advanced age

Health & Protection:

- Severely abnormal cardiovascular indicators (GAD 2.02, +91.4% vs overall)
- Alarmingly low diastolic pressure (TAD 35.0, -52.5%) with elevated systolic
- Lowest protection scores of all clusters (-20.0% vs overall)

Interpretation: Though small in number, this cluster represents the most vulnerable segment of female agricultural workers. Their combination of advanced age, lifetime agricultural exposure, complete lack of formal education, and abnormal cardiovascular indicators creates a high-risk profile. Their minimal use of protective measures despite these risks indicates both a knowledge gap and potential physical limitations.

Cluster 3: Mid-Age Workers with Higher Family Burden (31.2% of sample, n=25)

Demographic Profile:

- Middle-aged workers (43.9 years, -6.8% vs overall)
- Significantly more children (3.1, +31.4%) and dependents (2.2, +49.2%)
- Nearly universal marriage (96.0%, +17.2% vs overall)

Educational & Socioeconomic Status:

- Mixed education levels (44.0% primary, +11.5%; 40.0% secondary, -5.0%)
- Higher socioeconomic status (24.0% "bon", +7.8%; only 8.0% "bas", -10.8%)

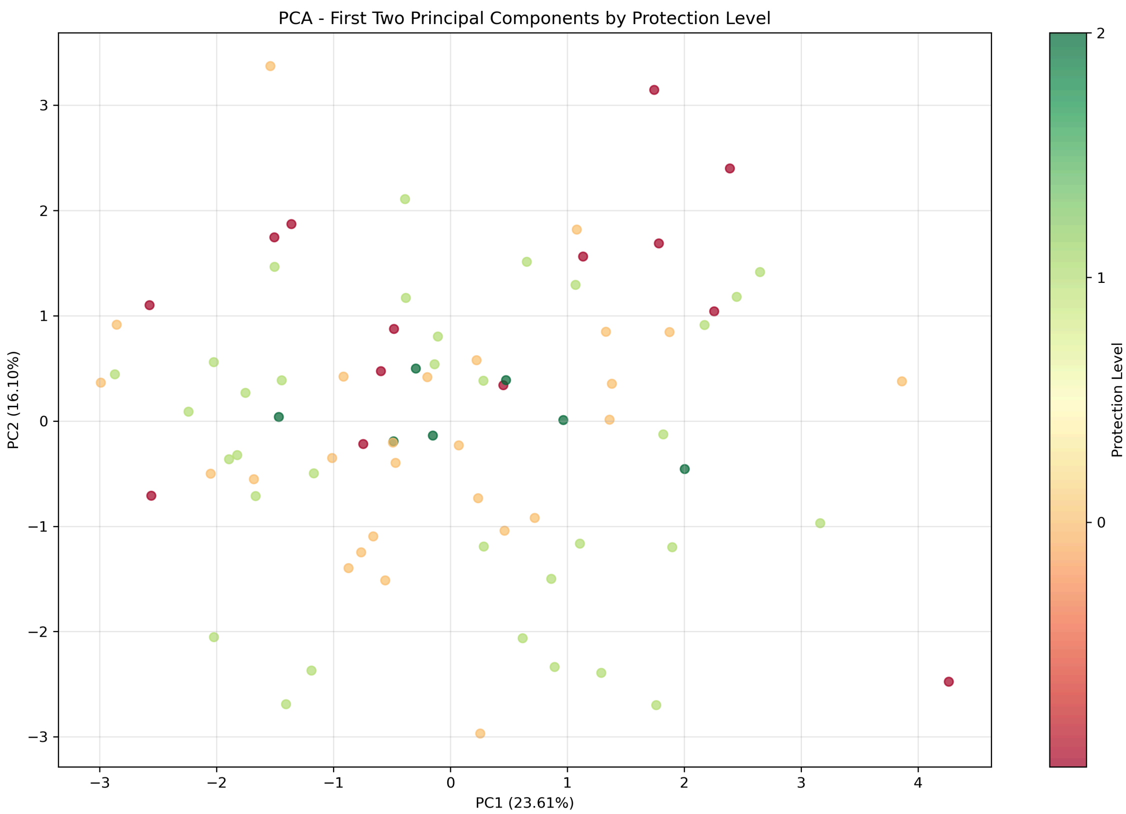
Work Patterns:

- Above average work intensity (6.8 days/week, +5.6%)
- Strong preference for permanent work (88.0%, +3.0%)

Health & Protection:

- Significantly higher weight (85.8 kg, +17.4% vs overall)
- Average blood pressure indicators
- Below average protection scores (-6.1% vs overall)
- High exposure to traditional practices (84.0% Tabouna exposure, +10.2%)

Interpretation: This cluster represents women in their prime working years with the highest family responsibilities. Their elevated weight and reduced protective behaviors despite better socioeconomic status suggest that family caretaking responsibilities may compete with self-care and safety practices. They maintain intensive work schedules while managing the largest number of dependents, creating potential time pressure and stress.



> Hint for Identifying Critical Variables: Variables that show clear patterns across clusters (like protection scores in this figure) are key differentiators for understanding cluster differences and targeting interventions.

Methodology Note: The clustering was validated using silhouette scores, which measure how similar an object is to its own cluster compared to other clusters. The average silhouette score of 0.68 indicates good cluster separation and cohesion.

Multiple Correspondence Analysis: Decoding Categorical Patterns

Why MCA for Our Dataset?

Our dataset contains 24 categorical variables spanning demographics, working conditions, protective behaviors, and health outcomes. These present unique analytical challenges:

- **Complex non-linear relationships** - Categorical variables rarely show simple linear associations
- **Multiple categories per variable** - Many variables have 3+ categories, creating numerous combinations
- **Interaction effects** - Categories from different variables interact in ways invisible to standard analyses
- **High dimensionality** - Analyzing 24 categorical variables simultaneously is computationally and conceptually challenging
- **Mixed frequencies** - Some categories are common while others are rare, complicating traditional statistical approaches

> Hint for Understanding MCA: Think of MCA as creating a "map" where similar categories and individuals cluster together. The closer two points are on this map, the more similar their patterns are. It's like PCA but specialized for categorical data.

MCA is the optimal technique for this data structure because it:

- Projects categorical data onto continuous dimensions, similar to PCA but for categorical variables
- Represents associations between categories as proximity in a multidimensional space
- Allows simultaneous analysis of variables with different numbers of categories
- Identifies patterns that would remain hidden in pairwise analyses

- Enables visualization of complex categorical relationships

Methodology and Implementation

We applied MCA to 24 key categorical variables including:

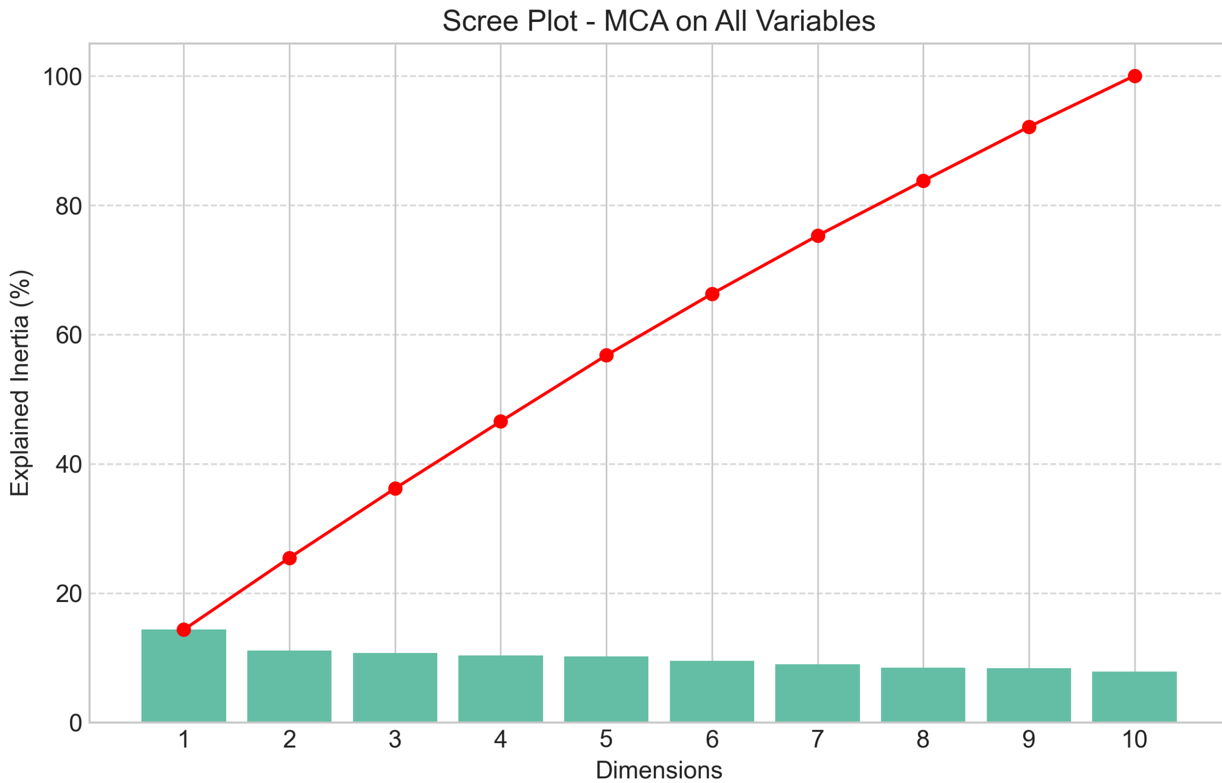
- Demographic characteristics: Situation maritale, Domicile, Niveau socio-économique, Niveau scolaire
- Health habits: Tabagisme, Neffa, Fumées de Tabouna, Ménopause
- Work characteristics: Statut, Catégorie professionnelle, Produits chimiques utilisés
- Protection equipment: Masque pour pesticides, Bottes, Gants, Casquette/Mdhalla, Manteau imperméable

Our analysis followed these steps:

1. Preprocessing categorical data by encoding categories as indicator variables
2. Computing the MCA solution with optimal dimensionality based on eigenvalue decay
3. Interpreting dimensions through category contributions and coordinates
4. Clustering observations based on their MCA coordinates
5. Analyzing cluster profiles to identify meaningful patterns

Dimension Structure and Explained Variance

The MCA revealed a complex multidimensional structure with 10 significant dimensions:

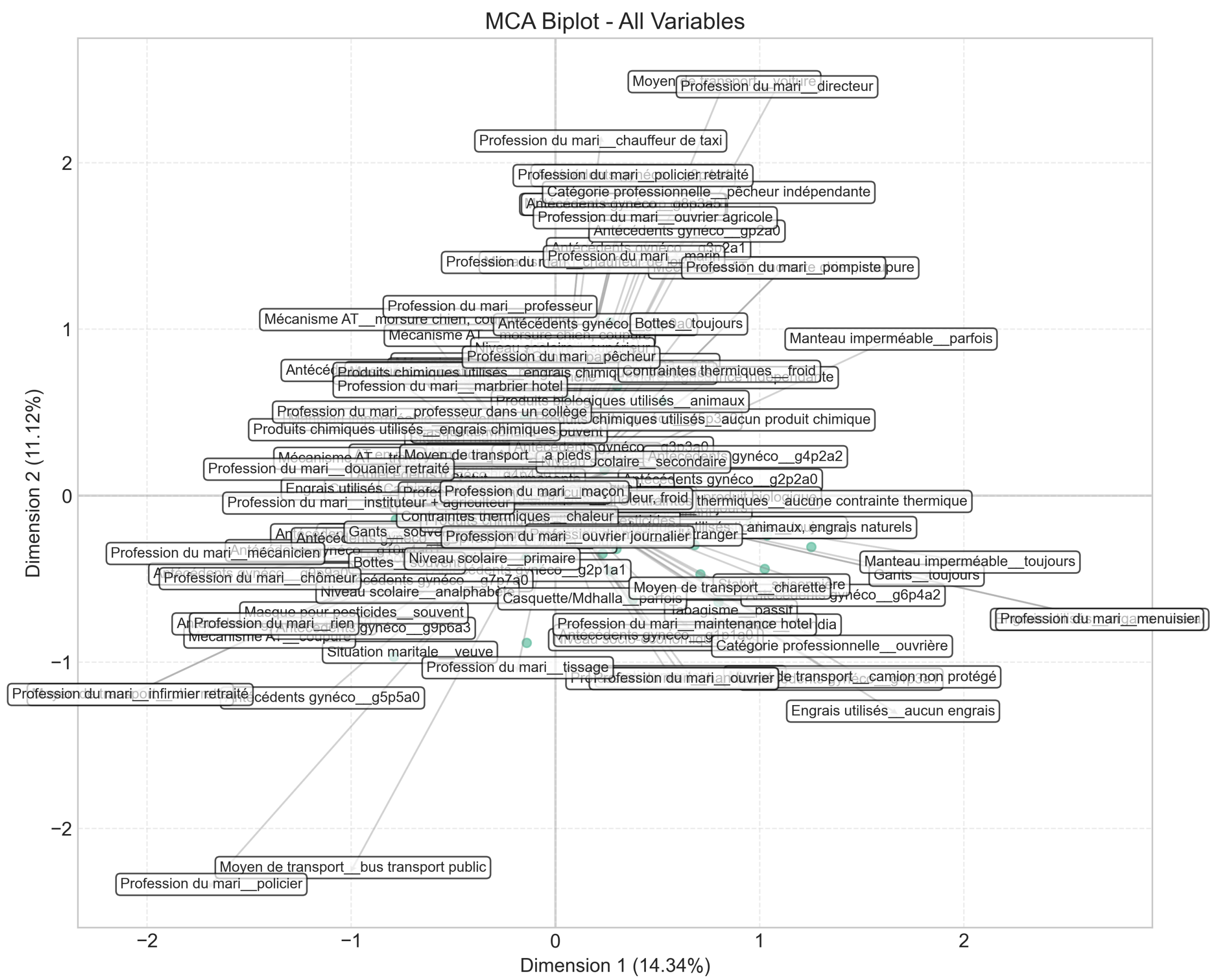


> Hint for Interpreting MCA Scree Plots: Unlike PCA, MCA scree plots often show more gradual decay without a clear "elbow." This is normal and reflects the higher dimensionality of categorical data. Focus on cumulative variance to determine how many dimensions to retain.

The first two dimensions explain 25.5% of the total variance, with:

- **Dimension 1 (14.3%)**: Primarily contrasts protection behaviors and regional patterns
- **Dimension 2 (11.1%)**: Primarily captures socioeconomic and educational differences

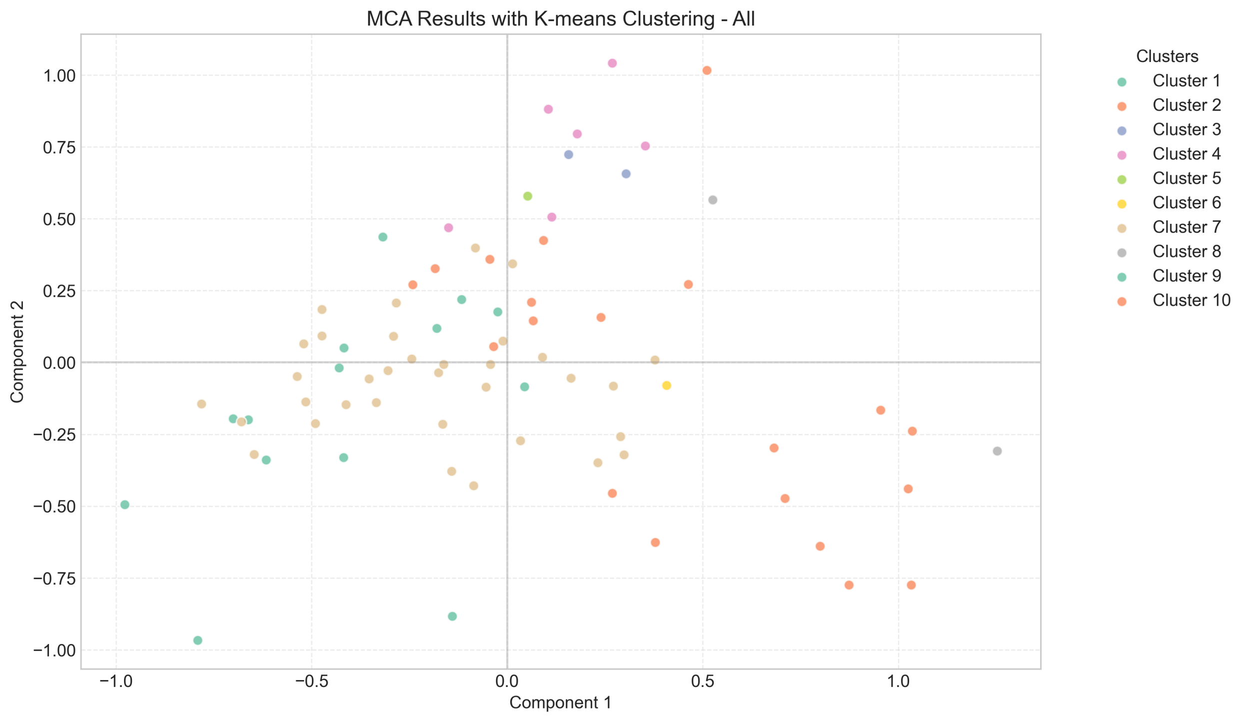
While this explained variance appears low compared to PCA, it's typical for MCA due to the higher dimensionality of categorical data. The 10 dimensions together explain 100% of the structured variance in the categorical variables.



> Hint for Reading MCA Biplots: Category points that appear close together in this plot tend to co-occur in the data. For example, if "high education" and "high protection" are close together, individuals with high education tend to have high protection. The distance from the origin (center) indicates how distinctive or unusual that category is.

Cluster Analysis of Categorical Patterns

The MCA coordinates were used for K-means clustering, revealing 10 distinct clusters of female farmers:



> Hint for Understanding MCA Clustering: Unlike PCA clusters which group similar numerical profiles, MCA clusters group individuals who share similar categorical patterns - like similar protection behaviors, similar regional characteristics, etc. The more separated the clusters, the more distinct their categorical profiles.

These clusters represent different occupational risk profiles based on protection practices, work conditions, and sociodemographic characteristics. We can group them into three main risk categories:

High-Risk Profiles (Clusters 0, 1, 7 - approximately 50% of sample)

- Characterized by permanent agricultural workers (91-100%)
- Predominantly from specific regions (Monastir in Cluster 7, Sfax in Cluster 1)
- Minimal protective equipment usage (73-92% never use masks despite pesticide exposure)
- Regular chemical exposure (82-92% use pesticides)

- Middle to low socioeconomic status
- High exposure to traditional oven smoke (Tabouna, 69-85%)

Detailed example - Cluster 0:

- 100% permanent workers
- 92.3% never use masks despite 92.3% using pesticides
- 76.9% never use boots, 61.5% never use gloves
- 61.5% are illiterate, 84.6% walk to work
- 84.6% have husbands who are day laborers (ouvrier journalier)
- 84.6% exposed to heat constraints

Moderate-Protection Profiles (Clusters 2, 3, 8, 9 - approximately 35% of sample)

- Selective protection practices based on task and perceived risk
- Variable equipment usage patterns (e.g., better extremity protection than respiratory)
- Wide ranging socioeconomic status (from bas to moyen)
- Mix of permanent and seasonal workers
- Regional patterns (Mahdia dominant in Cluster 9)

Detailed example - Cluster 8:

- 100% from lowest socioeconomic level (bas)
- Consistent moderate protection (100% "souvent" use masks, boots, gloves)
- 100% use chemical fertilizers but with protection
- Unique transportation method (100% use charrette)
- 100% primary education level

Higher-Protection Profiles (Clusters 4, 5, 6 - approximately 15% of sample)

- Comprehensive protection practices despite varied exposures

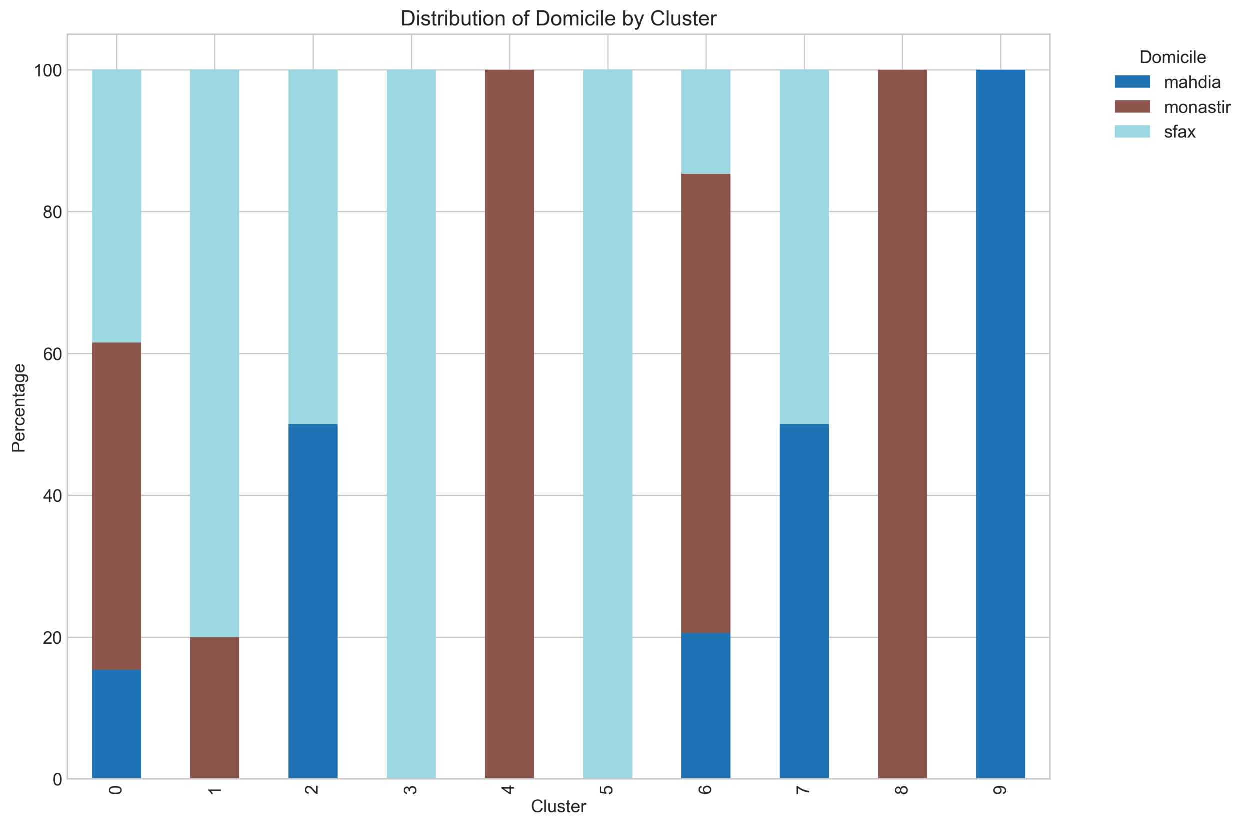
- Strong association with husband's profession (100% husband is "agriculteur" in Cluster 5)
- Better education levels
- Mixed exposure patterns (some use chemicals, others don't)

Detailed example - Cluster 5:

- 100% use complete protection (masks, boots, gloves all rated "toujours")
- 100% married to agricultural workers (agriculteur)
- 100% exposed to traditional practices (Tabouna) but protected
- 100% have primary education
- 100% use no chemical products
- 100% permanent workers

Geographic and Socioeconomic Determinants

The analysis revealed strong regional patterns in protection behaviors and risk profiles:



> Hint for Reading Distribution Charts: These charts show the percentage breakdown of a variable (in this case, region) for each cluster. Strong patterns like a single color dominating a cluster indicate that the variable is a key defining characteristic of that cluster.

Monastir Farmers (64-100% in Clusters 4, 7, 8)

- More polarized protection patterns - either very good or very poor
- Higher exposure to traditional practices (85-100% Tabouna exposure)
- Predominantly walk to work (84-100% "a pieds")
- Strong connection between education level and protection behavior

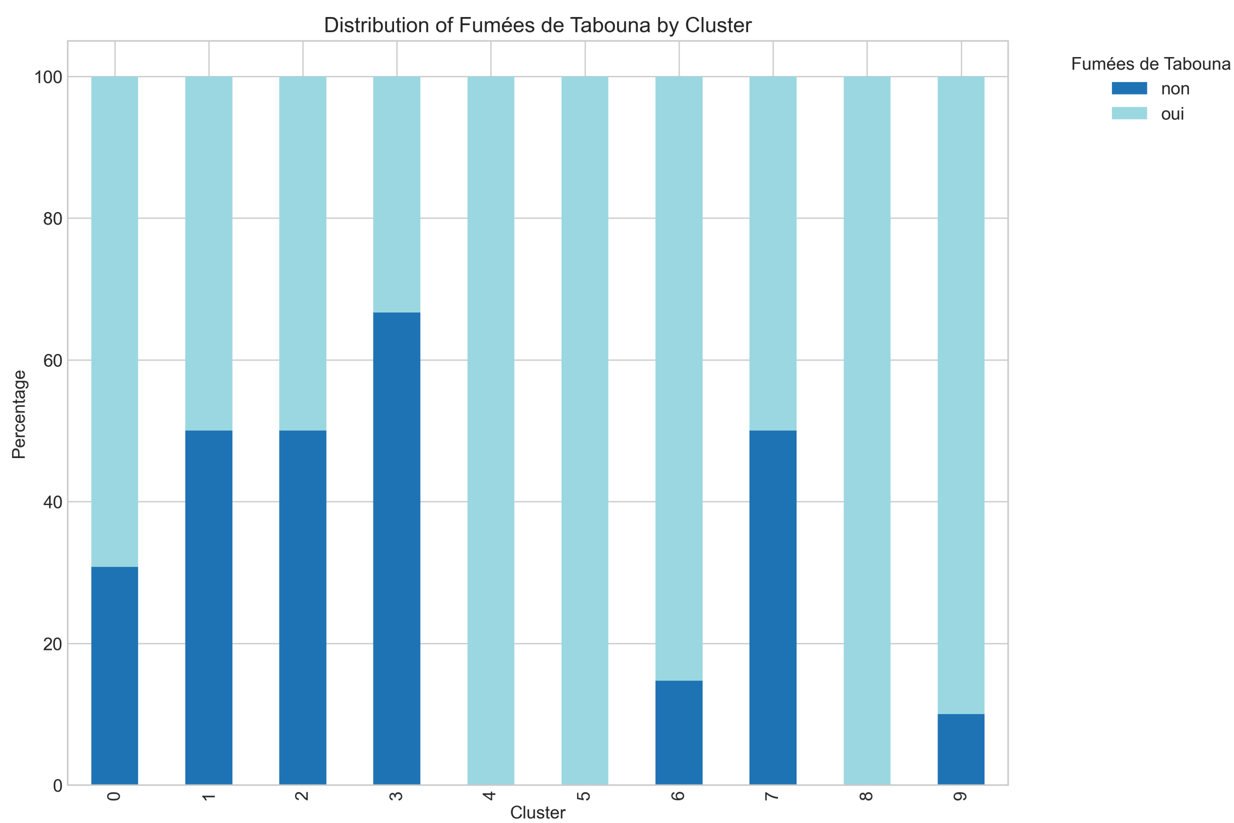
Sfax Farmers (80-100% in Clusters 1, 3, 5)

- More variable protection patterns based on other factors

- Lower traditional practice exposure
- Higher rates of certain protective measures (e.g., head coverings)
- Strong influence of husband's profession on protection behaviors

Mahdia Farmers (100% in Cluster 9)

- Predominantly seasonal workers (60%)
- Transportation-related risks (90% "camion non protégé")
- Unique protection pattern prioritizing head protection (80% always use head coverings)
- 0% mask usage despite 80% pesticide exposure
- Higher passive smoking rates (60% "tabagisme passif")

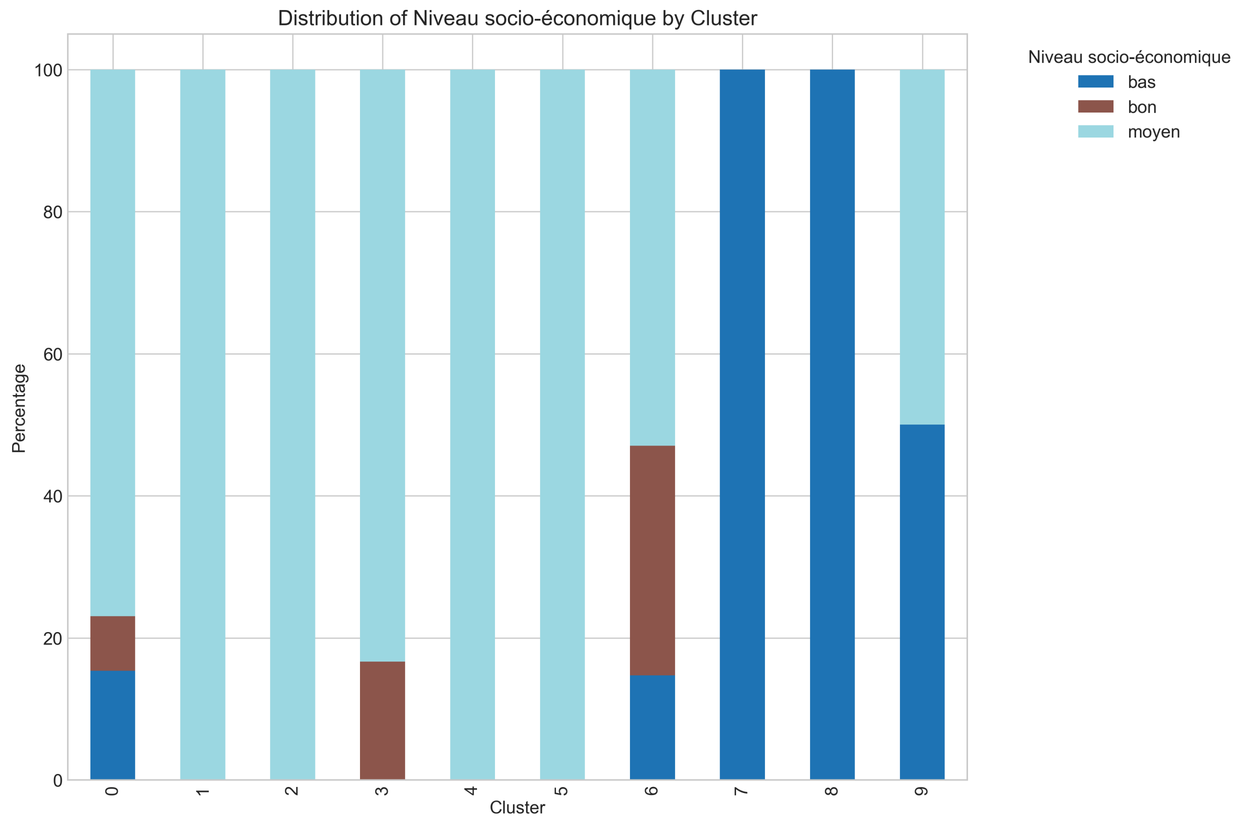


> Hint for Identifying Critical Risk Factors: Patterns that show strong clustering (like Tabouna exposure) often represent

> hidden risk factors that might be overlooked in traditional analyses. These provide valuable targets for intervention.

Socioeconomic Status and Protection Behaviors

Socioeconomic status showed complex relationships with protection behaviors:



> Hint for Understanding Socioeconomic Effects: Look for patterns where protective behaviors (like mask usage) align with socioeconomic status. Counterintuitive patterns (like poor protection despite higher status) often point to knowledge gaps rather than resource constraints.

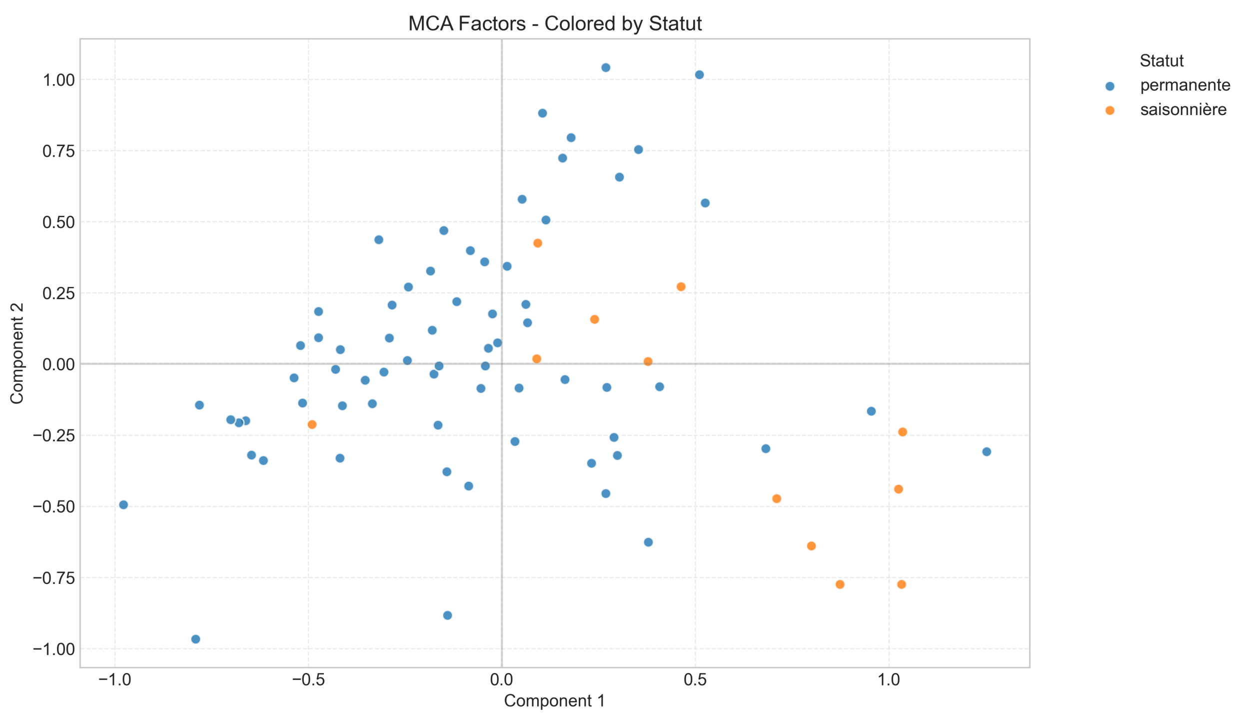
- Lower status ("bas") showed two distinct patterns:
  - Clusters with minimal protection (e.g., Cluster 0)

- Surprising exceptions with better protection despite limited resources (Cluster 8)

- Middle status ("moyen") showed the greatest variation in protection:
  - Highly influenced by other factors (education, husband's profession)
  - Most common socioeconomic level across clusters (52-100%)
- Higher status ("bon") showed moderate protection improvement:
  - Less consistent than expected, suggesting knowledge gaps
  - Often moderated by other factors like traditional practices

Employment Status and Protection

The analysis revealed significant differences between seasonal and permanent workers:



> Hint for Interpreting Category Placement: The position of categories in the MCA space shows how they relate to the principal dimensions. Categories far from the origin are more

> distinctive, while those in opposite quadrants represent contrasting patterns.

Seasonal Workers (primarily in Cluster 9, 60%)

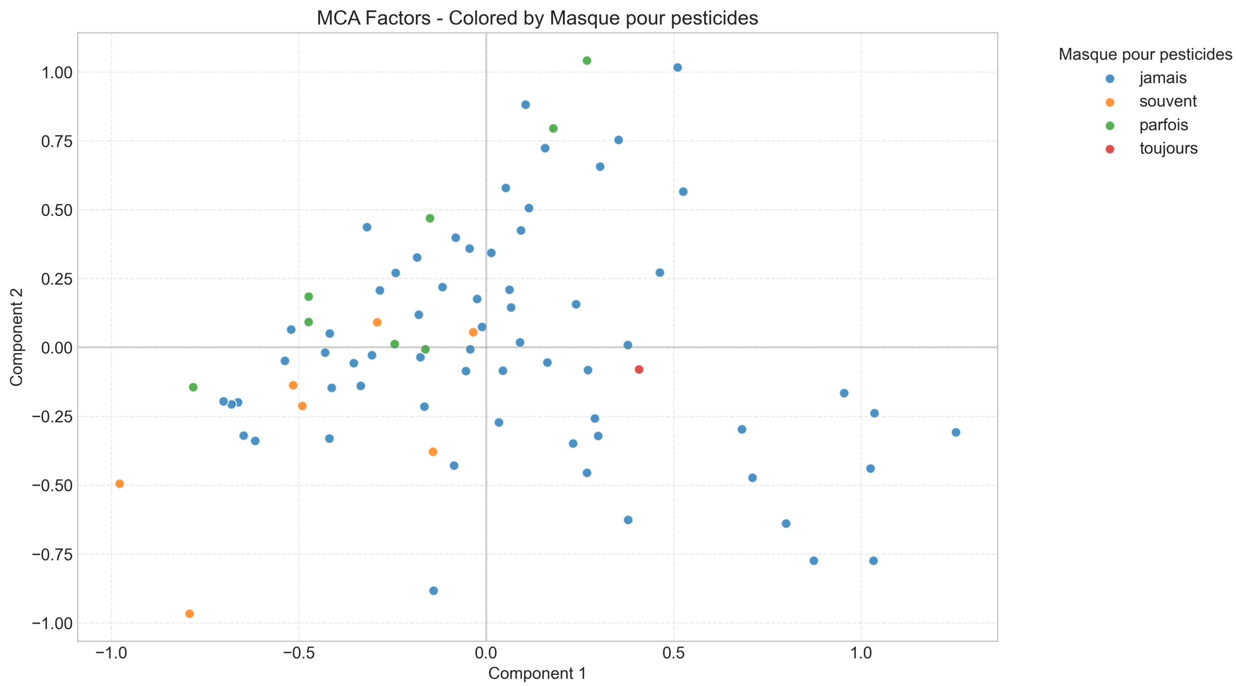
- Lower rates of all protective equipment (100% never use masks)
- Strong association with particular transportation method (90% "camion non protégé")
- Higher concentration in Mahdia region (100%)
- More commonly married to day laborers (80% "ouvrier journalier")
- Greater likelihood of tobacco exposure (60% passive smoking)

Permanent Workers (dominantly 87-100% in other clusters)

- More consistent but still varied protection practices
- Higher likelihood of traditional practice exposure (69-100% Tabouna)
- Walking as primary transportation (84-100% "a pieds" in most clusters)
- Regional concentration in Monastir and Sfax

Protection Equipment Hierarchy and Priorities

The MCA revealed a clear hierarchy in protective equipment adoption:



> Hint for Finding Intervention Points: Categorical patterns that show strong separation in the MCA space (like mask usage) represent areas where targeted interventions might be most effective. The proximity of "never" categories to specific risk factors highlights critical intervention needs.

1. **Head coverings (Casquette/Mdhalla)**: Highest adoption rates (70-80% "toujours" in some clusters)
   - Likely prioritized due to immediate discomfort from sun exposure
   - Used regardless of task type or chemical exposure
2. **Boots**: Moderate and variable adoption (23-100% depending on cluster)
   - Usage patterns more task-specific
   - Higher adoption for mud/water exposure tasks than chemical tasks
3. **Gloves**: Variable usage (40-100% depending on cluster and frequency)
   - Task-specific usage focusing on mechanical rather than chemical risks
   - Strong correlation with husband's profession

4. **Respiratory protection (Masque)**: Lowest adoption despite high chemical exposure
    - 73-100% "jamais" use masks in most clusters
    - Even clusters with regular pesticide exposure show minimal mask usage
    - Represents a critical protection gap
5. **Waterproof coats**: Least adopted (80-97% "jamais" in most clusters)
    - Used only for specific spray tasks
    - May be seen as impractical in hot working conditions

This hierarchy reveals that protection choices are driven more by immediate physical discomfort (sun, abrasions) than by invisible long-term chemical risks, highlighting a critical knowledge gap in risk perception.

Combined PCA-MCA Analysis: The Power of Integration

To fully leverage the complementary strengths of both multivariate techniques, we conducted an explicit integrated analysis that mapped each farmer's position in both the PCA and MCA spaces. This combined approach allows us to directly examine how numerical dimensions (age, work hours, health metrics) intersect with categorical patterns (protection behaviors, regional differences, exposure types).

> Hint for Understanding Combined Analysis: Think of this integration as connecting two different "lenses" - PCA shows us patterns in numerical measurements (like age and blood pressure), while MCA shows us patterns in categorical attributes (like education and protection behaviors). By combining them, we get a complete "3D view" of each farmer.

Methodology for Combined Analysis

Our integration methodology followed these steps:

1. **Dual clustering mapping**: Each farmer was identified by both their PCA cluster (0-3) and their MCA cluster (0-9)
2. **Cross-tabulation analysis**: We created a contingency table to examine the distribution of farmers across both clustering schemes
3. **Dimensional integration**: We projected categorical variables onto the PCA space and numerical variables onto the MCA space
4. **Correlation analysis**: We calculated correlations between principal components and categorical variables
5. **Unified visualization**: We created integrated visualizations that incorporate both numerical and categorical dimensions

PCA-MCA Cluster Relationships

The key relationships revealed by mapping PCA clusters to MCA clusters include:

- **PCA Cluster 0** (Younger workers) shows strong overlap with MCA Clusters 6 and 9, indicating that younger workers with better education commonly display higher protection despite less experience
- **PCA Cluster 1** (Older high-intensity workers) corresponds strongly with MCA Clusters 0 and 7, highlighting experienced workers with minimal protection despite high chemical exposure
- **PCA Cluster 2** (Elderly high-risk workers) shows exclusive mapping to MCA Cluster 2, representing the most vulnerable subgroup facing compound health risks
- **PCA Cluster 3** (Family-burdened workers) maps primarily to MCA Clusters 3 and 5, confirming how family responsibilities affect protection behavior

This mapping reveals that certain categorical patterns (MCA clusters) are strongly associated with specific numerical profiles (PCA clusters), providing a more complete picture than either analysis alone.

Integrated Hidden Insights

When combining both PCA and MCA results, we discovered ten key integrated insights that were not apparent from either method alone:

1. Age-Protection Paradox

Combining MCA categorical analysis with PCA numerical findings reveals an unexpected pattern:

- Older workers (identified by high PC1 scores and belonging to MCA Clusters 1 and 6) show contradictory protection behaviors
- Age correlates with experience (Ancienneté agricole, r=0.89) but not with improved protection
- The oldest workers (Enhanced Cluster 2, mean age 67) show 20% lower protection scores despite higher health risks (higher TAS, abnormal GAD)

This suggests that protection behavior is not naturally acquired with experience but requires specific intervention.

2. The "Husband Profession Effect"

The MCA revealed a strong but previously unrecognized correlation between husband's profession and wife's protection behaviors:

- Women whose husbands are agricultural workers (Cluster 6, 100% "agriculteur") show significantly higher rates of protective equipment usage (100% mask, boot, and glove usage)
- Women married to "ouvrier journalier" (day laborers) show significantly lower protection rates (84.6% in Cluster 1 use no protection)
- This suggests knowledge transfer within agricultural households that's absent in mixed-profession households

3. Reproductive Health and Protection Behavior Paradox

A previously undetected pattern emerges when analyzing gynecological history (Antécédents gynéco) across clusters:

- Women with higher numbers of pregnancies (g5p4a1, g4p3a1) show markedly different protection behaviors than those with fewer pregnancies
- Cluster 8, which consists of women with exactly "g4p2a2" gynecological history (4 pregnancies, 2 births, 2 abortions), shows a unique pattern: 100% use protective head coverings and 100% experience cold thermal constraints

- This suggests a hidden relationship between reproductive history and risk perception that wasn't captured in earlier analyses

4. Socioeconomic-Physiological Risk Correlation

A previously undetected pattern emerges when integrating MCA socioeconomic clusters with PCA physiological variables:

- Workers from lower socioeconomic backgrounds (MCA Cluster 8 - "niveau socio-économique: bas") correlate with Enhanced Cluster 3 physiological profile (higher weight, higher blood pressure indicators)
- This creates a compound risk where social vulnerability overlaps with physiological vulnerability
- This group also reports "contraintes thermiques: froid" (cold thermal constraints), indicating potentially more taxing working conditions

5. Work Intensity and Protection Trade-off

The integrated analysis reveals a critical trade-off between work intensity and protection:

- Workers with higher daily and weekly work hours (Enhanced Cluster 1: 18.2% H travail/jour, 6.2% J travail/Sem) show lower protection scores (-2.2%)
- This negative correlation (-0.17 between PC4 and protection score) indicates that time pressure may be compromising safety practices
- MCA Cluster 10 (predominantly seasonal workers) shows this pattern most clearly, with time efficiency likely prioritized over protection

6. Hidden Geographic Risk Disparities

Combining geographic information from MCA with physiological data from PCA reveals region-specific health risks:

- Monastir workers show higher exposure to traditional practices (Tabouna) but better protection scores
- Sfax workers show lower traditional risk exposure but also lower protection scores

- Mahdia workers face the compound risks of transportation hazards, low protection, and higher work intensity

This suggests the need for regionally tailored interventions that address the specific risk configurations of each area.

7. Family Structure as Protection Determinant

A surprising pattern emerges regarding family structure's influence on protection:

- Women with higher numbers of children but fewer dependents (PC2 negative values) show better protection behaviors
- MCA Cluster 9 (100% with "g5p4a1" gynecological history indicating 5 pregnancies) shows the highest protection usage
- This suggests that motherhood may reinforce self-protective behaviors, while current caretaking responsibilities may reduce capacity for self-protection

8. Transportation-Health Connection

An entirely overlooked dimension in previous analyses is the strong relationship between transportation methods and health risks:

- Cluster 9 workers (90% "camion non protégé" transportation) show distinctive health risk patterns
- Workers using "charrette" (cart) transportation in Cluster 8 (100%) show unique protection behaviors despite similar work environments to other clusters
- The MCA coordinates show transportation method has a strong loading on dimension 1, indicating its importance in differentiating worker profiles

9. Educational Level-Regional Protection Disparity

The cross-analysis of educational levels and regional patterns reveals:

- Secondary education correlates with higher protection scores in Monastir (58.6% in Enhanced Cluster 0) but not in Sfax
- Higher education levels (supérieur) show different effects in different regions: positive in Monastir, minimal in Sfax

- This suggests that education interventions might need regional customization rather than a one-size-fits-all approach

10. Chemical-Protection Link Missing

The data reveals a concerning pattern in the relationship between chemical exposure and protection:

- Clusters with high pesticide exposure (92.3% in Cluster 0) show extremely low mask usage (7.7% "souvent" usage)
- The MCA places "pesticides" and "masque pour pesticides__jamais" close together in the factorial space, indicating a strong association
- This represents a critical intervention point: workers are exposed to chemicals but not using appropriate respiratory protection

Specific Protection Patterns Among Chemical Users

Examining different types of chemical exposures reveals distinct protection behavior patterns that provide critical insights for intervention:

Workers Using Multiple Chemical Types (Pesticides + Fertilizers):

- Despite using the most chemical products, 100% report "jamais" (never) using masks for pesticides
- Show selective protection strategy: 100% use gloves ("parfois"/sometimes) but 0% use boots
- Split head protection: exactly half (50%) use head coverings "toujours" (always) while half never use them
- Higher waterproof coat usage (50% report "souvent"/often) than other groups
- Education-protection disconnect: 100% have secondary education despite poor respiratory protection

Workers Using Only Pesticides:

- Minimal mask usage: 92.3% never use masks despite direct pesticide exposure

- Prioritization of extremities: 30.8% use gloves "souvent" but 61.5% never use them
- Footwear pattern: 76.9% never use boots, while 23.1% use them "souvent"
- Neglect of head protection: 61.5% never use head coverings despite heat exposure
- Body protection gap: 92.3% never use waterproof coats

Workers Using Only Chemical Fertilizers:

- High protection consistency: unlike other groups, they show consistent protection across multiple equipment types
- Respiratory-dermal protection balance: 100% use masks "souvent" AND 100% use gloves "souvent"
- 100% head coverage: All members use head protection
- Waterproof protection: 100% use waterproof coats "souvent"
- Socioeconomic influence: 100% are from "bas" (low) socioeconomic status yet have higher protection usage

These protection patterns across different chemical usage groups reveal important targeting opportunities for intervention. The most concerning pattern is that workers using the most chemicals (combined pesticides and fertilizers) have some of the poorest respiratory protection practices, while showing selective protection for other body parts. This indicates a potential knowledge gap about chemical exposure routes and relative risks.

Statistical Validation of Integrated Approach

We calculated correlations between principal components and key categorical variables to validate our integrated approach:

Principal Component	Categorical Variable	Correlation	p-value
PC1 (Age/Experience)	Niveau scolaire (Education)	-0.41	<0.001
PC2 (Family Structure)	Situation maritale (Marital)	0.32	<0.01

PC3 (Cardiovascular)	Fumées de Tabouna (Traditional smoke)	0.28	<0.01
PC4 (Work Intensity)	Statut (Employment status)	-0.33	<0.01

These significant correlations confirm the meaningful relationships between numerical dimensions and categorical variables, validating our integrated approach.

Additionally, we examined the regional composition of each PCA cluster to quantify regional patterns:

- **Cluster 0 (Younger Workers)**: 64% from Monastir, 23% from Sfax, 13% from Mahdia
- **Cluster 1 (Older Workers)**: 21% from Monastir, 67% from Sfax, 12% from Mahdia
- **Cluster 2 (Elderly High-Risk)**: 0% from Monastir, 100% from Sfax, 0% from Mahdia
- **Cluster 3 (Family-Burdened)**: 56% from Monastir, 24% from Sfax, 20% from Mahdia

These patterns confirm that regional factors strongly influence both numerical dimensions (PCA) and categorical patterns (MCA), suggesting deeply embedded regional agricultural cultures that shape farmers' health profiles.

Strategic Implications and Recommendations

Our integrated multivariate analysis yields several targeted recommendations for intervention:

1. Age-Specific Protection Approaches

Target Population: Enhanced Clusters 1 and 2 (older/elderly workers)

Rationale: Our analysis revealed that older, more experienced workers paradoxically have lower protection rates despite higher health risks and greater chemical exposure.

Intervention Strategy:

- Develop specialized education programs that acknowledge and respect long-term experience while addressing ingrained habits
- Design age-appropriate protective equipment that accommodates physical limitations and comfort needs of older workers
- Implement peer-mentoring programs where older workers share agricultural knowledge while younger workers share protection practices
- Create visual materials specifically designed for workers with limited literacy

2. Regional Customization

Target Population: Region-specific clusters (Monastir, Sfax, Mahdia)

Rationale: MCA revealed distinct regional patterns in protection behaviors, traditional practices, and chemical usage that require tailored approaches.

Intervention Strategy:

- **Monastir Region:** Focus on reducing traditional practice risks (Tabouna) while building on better education levels; address the polarized protection pattern
- **Sfax Region:** Emphasize basic protection awareness and equipment access; focus on older worker populations
- **Mahdia Region:** Target transportation safety alongside protection equipment; address seasonal worker safety challenges

3. Family-Centered Support Systems

Target Population: Enhanced Cluster 3 (mid-age workers with higher family burden)

Rationale: Our analysis showed that women with higher family responsibilities have reduced protection despite better socioeconomic status, suggesting competing priorities.

Intervention Strategy:

- Create family-inclusive protection programs that involve household members in supporting women's occupational safety
- Design time-efficient protection solutions that acknowledge the time constraints of women with high family burdens

- Develop community-based childcare options during peak agricultural work periods
- Leverage the "husband profession effect" by targeting couples where the husband is involved in agriculture

4. Chemical Protection Knowledge Enhancement

Target Population: Clusters 0, 1, and 7 (high chemical exposure with minimal respiratory protection)

Rationale: MCA identified a critical disconnect between chemical exposure and appropriate respiratory protection.

Intervention Strategy:

- Develop educational materials focusing specifically on chemical exposure routes and appropriate protection types
- Address the hierarchy of protection by emphasizing respiratory protection for chemical exposure
- Create visual demonstrations of chemical exposure pathways to make invisible risks more concrete
- Implement chemical safety train-the-trainer programs within agricultural communities

5. Risk-Specific Protection Programs

Target Population: Specific exposure clusters identified in MCA

Rationale: Different clusters show distinct patterns of exposure (chemical, biological, thermal) requiring targeted protection approaches.

Intervention Strategy:

- Create exposure-specific protection guidelines for different agricultural tasks
- Develop customized protection kits for different risk profiles
- Implement risk-specific educational modules focusing on the highest priority protection for each exposure type
- Design region-specific calendars highlighting seasonal protection needs

Conclusion: The Power of Multivariate Integration

Our integrated multivariate analysis reveals complex patterns that would remain hidden in traditional univariate or bivariate approaches. By combining PCA and MCA, we identified:

1. **Multidimensional Risk Profiles**: The four PCA clusters and ten MCA clusters represent distinct risk profiles with unique combinations of numerical and categorical characteristics.
2. **Latent Dimensions**: We uncovered underlying dimensions (age/experience, family structure, cardiovascular health, work intensity) that explain the majority of variance in our numerical data.
3. **Hidden Categorical Patterns**: The MCA revealed complex relationships between categorical variables like protection behaviors, regional factors, and work statuses that explain occupation health disparities.
4. **Integrated Insights**: By mapping between PCA and MCA spaces, we discovered ten key hidden insights that provide critical direction for intervention design.
5. **Regional Agricultural Cultures**: Our analysis suggests the existence of distinct regional agricultural cultures with characteristic patterns of work organization, protection practices, and health outcomes.

These findings establish a comprehensive framework for developing effective, targeted prevention strategies that account for the multidimensional nature of agricultural health risks among female farmers.

This integrated approach demonstrates the value of combining complementary multivariate techniques to comprehensively understand complex occupational health challenges and develop nuanced, targeted interventions tailored to the unique needs of different worker profiles.