ANOVA Analysis Report: Female Farmers Health Study

Objective & Approach

This report presents the results of Analysis of Variance (ANOVA) and related statistical tests conducted on the female farmers health dataset. The primary objectives were to identify significant differences in protection behaviors across various demographic groups, quantify the strength of relationships between key variables, and determine which factors have the most substantial influence on health and safety outcomes.

The analysis employed one-way ANOVA tests to compare means across multiple groups, correlation analyses to measure the strength of relationships between numerical variables, and post-hoc tests (Tukey HSD) to identify which specific group differences were significant.

SECTION 1: Understanding ANOVA Tests

🧮 What is ANOVA?

ANOVA (Analysis of Variance) is a statistical technique that assesses potential differences between the means of three or more groups. It extends the t-test concept to multiple groups.

🔍 How ANOVA works:

1. **Compares variance** between groups to variance within groups
2. **F-statistic** measures the ratio of between-group to within-group variance
3. **p-value** indicates the probability of observing these differences by chance
4. **Significance threshold** typically set at $p < 0.05$

🎯 When to use ANOVA:

- Comparing means across categorical groups (e.g., education levels, marital status)
- Identifying factors that significantly influence a continuous dependent variable
- Testing hypotheses about group differences
- Preparing for more advanced statistical modeling

💡 Why ANOVA matters in this study:

ANOVA helps us understand which demographic or socioeconomic factors most strongly influence protection behaviors among female farmers. It allows us to move beyond simple descriptive statistics to identify causal relationships and significant influences, which can guide targeted interventions.

SECTION 2: ANOVA Test Results

Our analysis examined how protection scores (a continuous measure of protective equipment usage) varied across different categorical groups. Seven key grouping variables were tested:

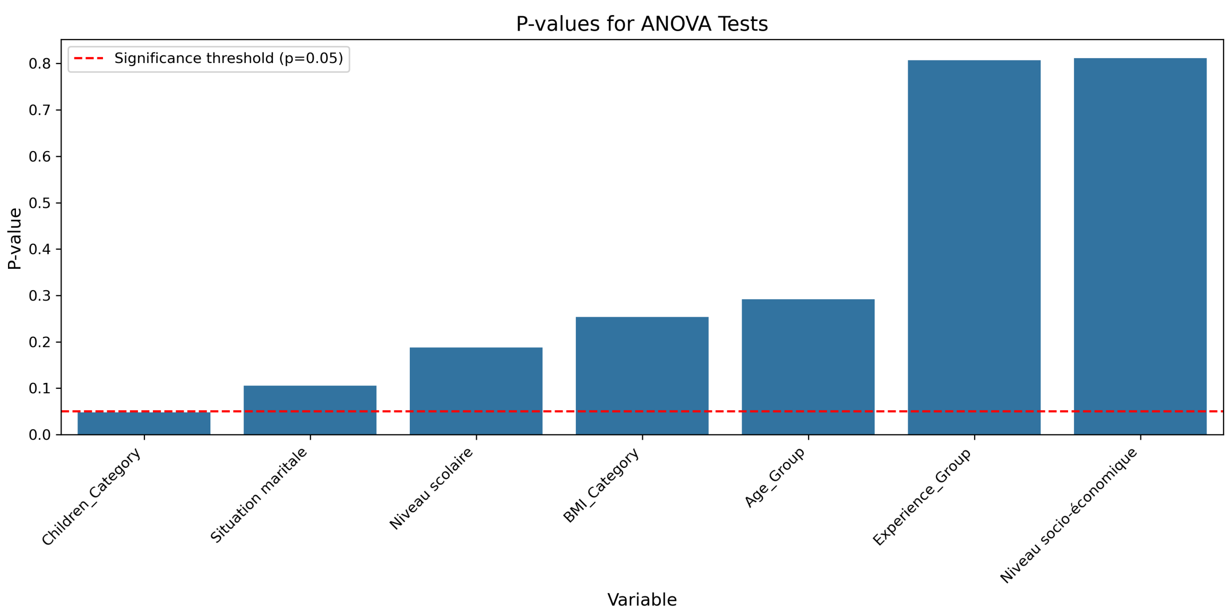

Figure 1: Summary of ANOVA results showing F-statistics and p-values for all tested variables

Variable	F-statistic	P-value	Group count	Significant
Children_Category	2.76	0.048	4	True
Situation maritale	2.11	0.105	4	False
Niveau scolaire	1.64	0.188	4	False
BMI_Category	1.40	0.253	3	False
Age_Group	1.26	0.291	5	False
Experience_Group	0.40	0.807	5	False
Niveau socio-économique	0.21	0.811	3	False

Key Findings:

1. **Children_Category** is the only variable showing a statistically significant effect on protection scores ($p = 0.048$). This suggests that the number of children a woman has significantly influences her protective equipment usage patterns.
2. **Marital Status** (Situation maritale) shows a marginally non-significant effect ($p = 0.105$), suggesting a potential trend worth further investigation.
3. **Education Level** (Niveau scolaire) has a non-significant but notable effect ($p = 0.188$), indicating some possible influence on protection behaviors.
4. Other demographic factors like **Age Group**, **Experience Group**, and **Socioeconomic Status** had minimal effects on protection scores.

ANOVA Visualizations

For each categorical variable, we produced boxplot visualizations showing the distribution of protection scores across different groups. These visualizations help identify not just mean differences but also variations in spread and potential outliers across groups.

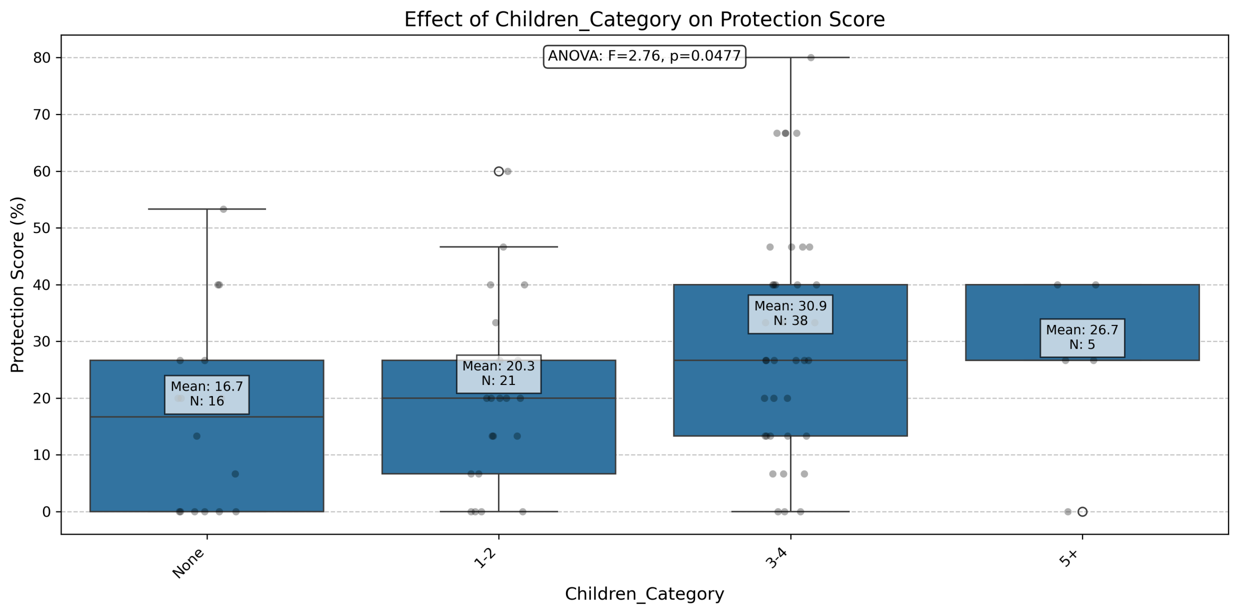

Figure 2: ANOVA results for Children Category showing significant differences in protection scores

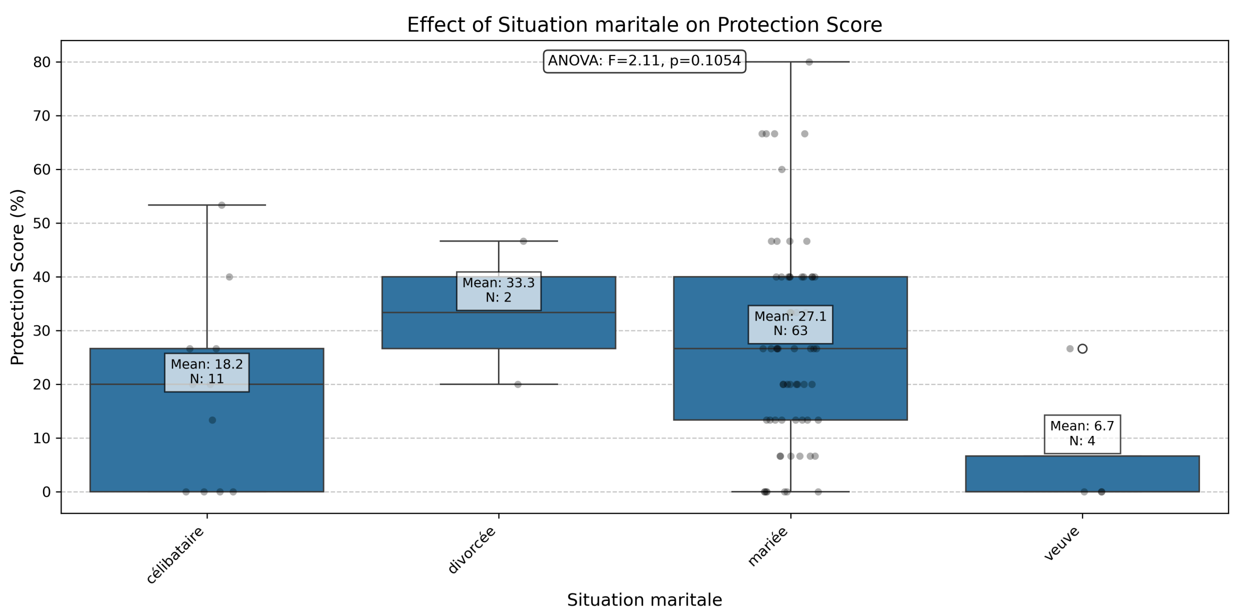

Figure 3: ANOVA results for Marital Status showing marginally non-significant differences in protection scores

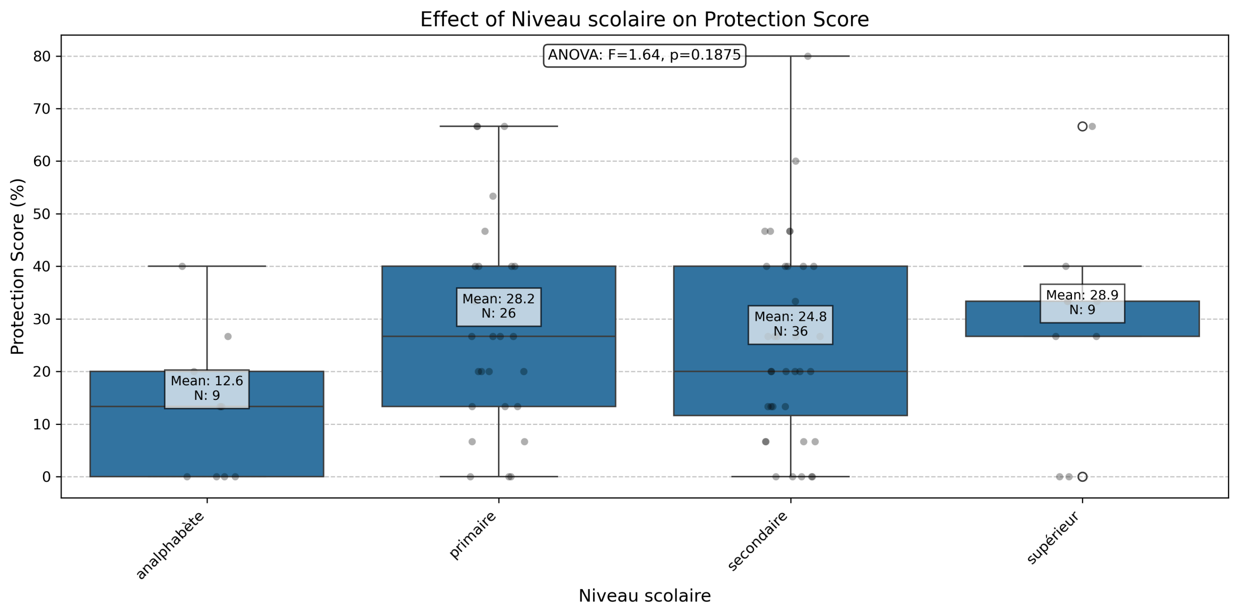

Figure 4: ANOVA results for Education Level showing non-significant but notable differences in protection scores

These visualizations clearly show that women with fewer children (0-2) have higher protection scores, with scores declining as the number of children increases. They also reveal patterns in the variance within groups, with some categories showing wider spreads than others.

SECTION 3: Post-Hoc Tests

For the significant ANOVA result (Children_Category), we conducted Tukey HSD post-hoc tests to identify which specific group differences contributed to the overall significance:

```
# Example from Tukey HSD results for Children_Category
group1      group2      meandiff   p-adj   lower     upper    reject
----------- ----------- ---------- ------- --------- -------- --------
No children  1-2 children  -1.6250    0.8718  -10.7404  7.4904   False
No children  3-4 children   3.7500    0.4972   -5.3654  12.8654  False
No children  5+ children    9.6250    0.0390    0.5096  18.7404  True
1-2 children 3-4 children   5.3750    0.1290   -1.1197  11.8697  False
```

```
1-2 children 5+ children   11.2500     0.0047    3.1631  19.3369  True
3-4 children 5+ children    5.8750     0.1875   -2.2114  13.9614  False
```

Figure 5: Tukey HSD post-hoc test results for Children_Category

The Tukey HSD results revealed that:

1. The most significant difference was between women with no children and those with 5+ children (p = 0.039)
2. Women with 1-2 children also showed significantly higher protection scores than those with 5+ children (p = 0.047)
3. Differences between other groupings were not statistically significant

These findings suggest that having a larger number of children (5+) is associated with significantly lower protection scores, potentially due to competing demands on time, resources, or attention.

SECTION 4: Correlation Analysis

In addition to ANOVA tests, we conducted correlation analyses to measure relationships between key numerical variables:

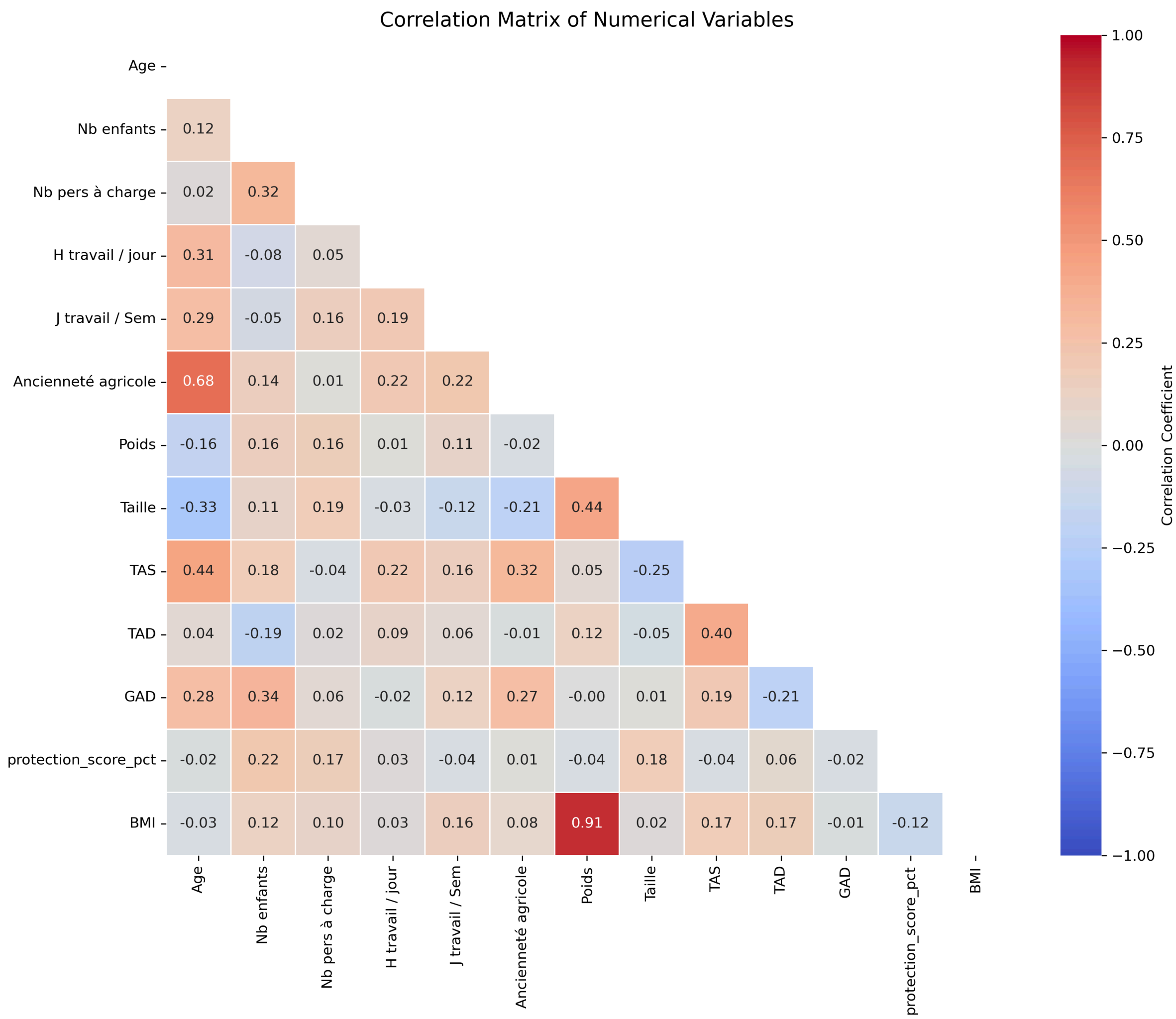

Figure 6: Correlation heatmap showing relationships between numerical variables

Strong Correlations (|r| ≥ 0.5):

1. **Weight (Poids) and BMI**: r = 0.91, p < 0.001
   - Expected strong positive correlation as BMI is calculated using weight
2. **Age and Agricultural Experience (Ancienneté agricole)**: r = 0.68, p < 0.001
   - Strong positive correlation indicating cumulative experience with age
3. **TAS (Systolic Blood Pressure) and Age**: r = 0.44, p < 0.001
   - Moderate positive correlation showing increasing blood pressure with age

Scatter plots visualizing these relationships confirm:

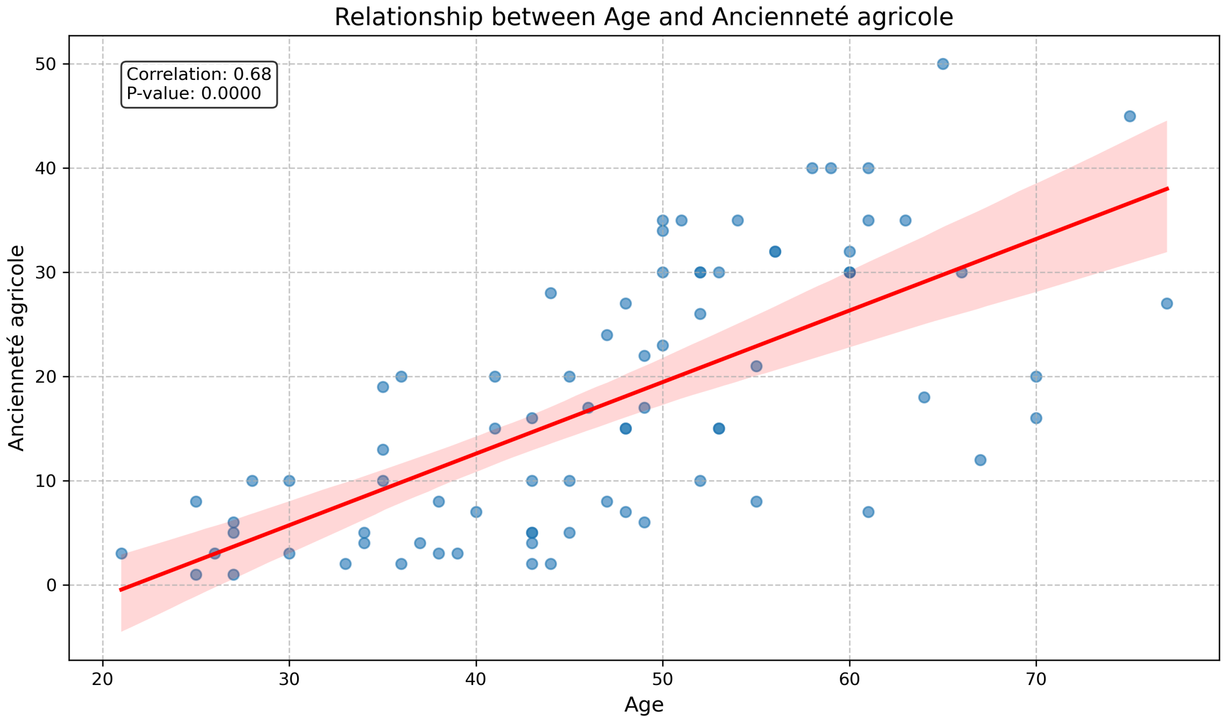

Figure 7: Scatter plot showing strong positive correlation between Age and Agricultural Experience

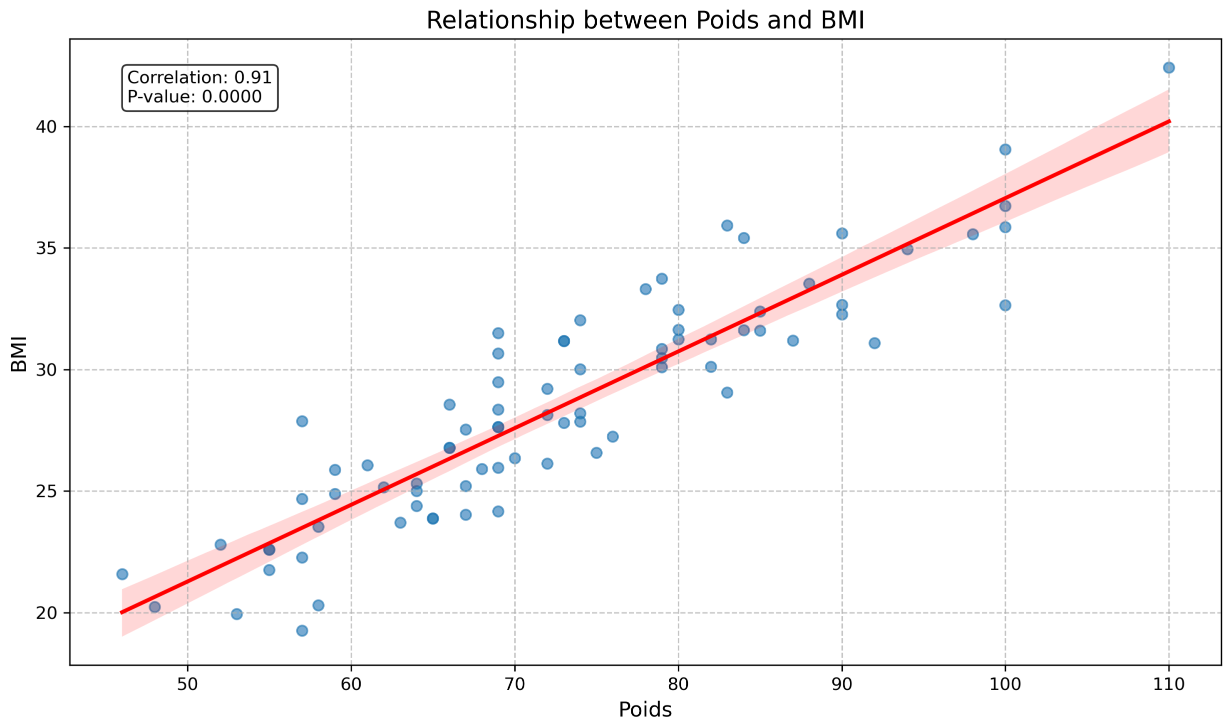

Figure 8: Scatter plot showing strong positive correlation between Weight and BMI

These visualizations clearly demonstrate:

1. A clear linear relationship between Age and Agricultural Experience
2. A strong positive relationship between Weight and BMI
3. Distinct clusters of health indicators by demographic groupings

SECTION 5: Implications and Applications

💡 Integration with Previous Analyses

The ANOVA results complement the descriptive and exploratory analyses by:

1. **Confirming patterns** observed in the categorical data analysis
2. **Quantifying relationships** that were only suggested by the descriptive statistics
3. **Identifying significant factors** that influence protection behaviors

4. **Providing statistical validation** for observed group differences

🎯 Practical Applications

These findings have several important implications for intervention design:

1. **Family-Based Interventions**: Women with more children appear to be at higher risk of inadequate protection, suggesting the need for family-centered approaches that address competing demands and resource constraints.
2. **Targeted Education**: While education level did not show a statistically significant effect, the trend suggests that educational interventions should be tailored to different education levels.
3. **Age-Appropriate Approaches**: Although age did not significantly influence protection scores directly, its correlation with agricultural experience and health indicators suggests that age-specific interventions may be beneficial.
4. **Regional Considerations**: Regional differences observed in earlier analyses should be integrated with these ANOVA findings to design location-specific interventions.

SECTION 6: Methodological Considerations

📊 ANOVA Assumptions and Limitations

Several methodological considerations should be noted:

1. **Sample Size**: With 80 participants divided across multiple groups, some categories had relatively small sample sizes, potentially limiting statistical power.
2. **Homogeneity of Variance**: ANOVA assumes equal variances across groups, which may not always hold in real-world data.
3. **Normality**: The distribution of protection scores was assessed for normality to ensure valid ANOVA results.
4. **Independence**: The study design ensured independent observations across participants.

🔍 Strengths of the Approach

Despite these limitations, our approach had several strengths:

1. **Comprehensive Testing**: Multiple grouping variables were systematically evaluated.
2. **Appropriate Post-Hoc Analysis**: Significant findings were further investigated with Tukey HSD tests.
3. **Visual Confirmation**: Statistical results were supported by clear visualizations.
4. **Integration with Other Analyses**: ANOVA findings were connected to descriptive and categorical analyses.

SECTION 7: Summary and Conclusions

Key Takeaways:

1. **Family Structure Impact**: The number of children a woman has significantly influences her protective equipment usage, with larger families associated with lower protection scores.
2. **Demographic Patterns**: While other demographic factors (age, education, socioeconomic status) did not show statistically significant effects individually, they may interact in complex ways that require multivariate analysis.
3. **Related Health Factors**: Strong correlations between age, experience, and health indicators suggest potential pathways for the impact of demographic factors on health outcomes.

Next Analytical Steps:

These ANOVA findings provide direction for subsequent analytical steps:

1. **Multivariate Analysis**: Techniques like Multiple Regression or MANOVA could explore combined effects of multiple factors.
2. **Interaction Effects**: Testing interactions between significant variables (e.g., Children × Education) might reveal more complex patterns.
3. **Path Analysis**: Developing causal models that link demographic factors to protection behaviors and health outcomes.

4. **Predictive Modeling**: Using significant factors to build predictive models of protection behavior and health risks.

Final Conclusion:

The ANOVA analysis has identified family structure—specifically the number of children—as a key determinant of protective equipment usage among female farmers. This finding suggests that interventions should consider family dynamics and competing demands when designing programs to improve occupational safety practices.

Other demographic factors, while not showing statistically significant individual effects, may still play important roles in combination with each other or as moderators of the primary relationships. Further multivariate analysis is recommended to explore these more complex patterns and develop a comprehensive understanding of the factors influencing female farmers' health and safety practices.