**The Prairie Project: Evaluating Rangeland Conservation Awareness Across Age Groups**

Sai Sushma Mutyala

Texas A & M University, STAT 692

02/14/2025

**Project Description**

This project involves a qualitative survey designed to assess the awareness and learning gains of different age groups (14-18, 19-40, and 41+) regarding rangeland management, prescribed fire, multi-species grazing, and conservation practices.

The study explores how prior experiences with these topics influence participants' knowledge and willingness to learn more about rangelands. By analyzing responses to targeted survey questions, this research provides insights into the effectiveness of conservation education and helps develop strategies to better engage different age groups in rangeland stewardship.

**Project objectives:**

This study aims to:

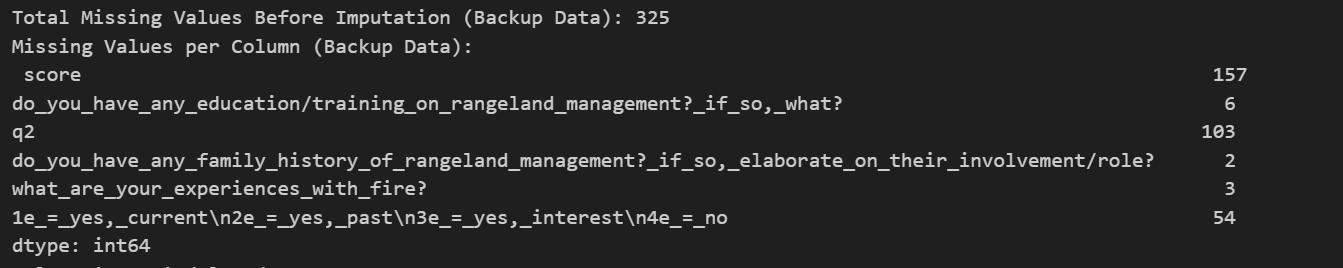
1. Assess awareness and learning gains across different age groups.
2. Examine whether prior experience with rangeland conservation influences knowledge acquisition.
3. Investigate the relationship between motivation (desire to learn) and learning outcomes.
4. Identify effective strategies for conservation education.

**Data Pre-processing:**

This project initially contained four separate datasets, with survey questions and responses split by age group (14-18, 19-40, and 41+). To facilitate analysis, these datasets were merged into a single structured dataset (df\_combined), ensuring that responses from all age groups were aligned with the corresponding survey questions. An "age\_group" column was added to track respondents' age categories. To maintain consistency, column names were standardized by converting them to lowercase and replacing spaces with underscores. The datasets were then merged, the index was reset, and a backup copy (df\_backup) was created to preserve the original data. The final dataset consists of 157 rows and 44 columns.

Before analysis, missing values were identified and addressed. The dataset initially contained 325 missing values across multiple columns, including the score column, education or training on rangeland management, family history of rangeland management, fire experience, and survey question q2. The score column, which was entirely missing and irrelevant to the analysis, was removed. Mode imputation (replacing missing values with the most frequent category) was applied within each age group to preserve response patterns and minimize bias. After imputation, all missing values were successfully addressed, and no duplicates were found in the dataset.

The dataset is now cleaned, structured, and ready for further analysis.



**Methods:**

### 4.1 Hypotheses and Model Selection

This study examines whether prior experience and age group influence grassland conservation learning gains and perceptions about prescribed burns. The hypotheses tested are:

#### Client-Provided Hypotheses

1. Does prior experience with WPE impact grassland conservation learning gains across age groups?
2. Does prior experience with fire impact grassland conservation learning gains across age groups?
3. Does prior experience with patch burn grazing impact learning gains across age groups?
4. Does prior experience with multi-species grazing impact learning gains across age groups?
5. Does prior experience with rangeland management increase the desire to learn more about rangelands in school?

#### Additional Hypothesis

1. Does age group influence opinions on prescribed burns?

To test these hypotheses, categorical data analysis methods were selected. Since both independent variables (prior experience and age group) and dependent variables (learning gains and opinions) are categorical, the most appropriate statistical method is the Chi-Square Test for Independence.

The Chi-Square test is used to determine if there is an association between prior experience and learning gains or between age group and prescribed burn opinions. If the Chi-Square test assumptions are violated due to small sample sizes, Fisher’s Exact Test is applied.

Machine learning models such as logistic regression or decision trees were not chosen because they require numerical transformations of categorical data, which may introduce bias, and the dataset size is insufficient for predictive modeling.

#### Contingency Table Setup

For learning gains hypotheses:

* Independent Variable: Prior experience (Yes/No)
* Dependent Variable: Learning gains (Yes/No)
* Age Groups: 14-18, 19-40, 41+
* 2 × 2 × 3 table

For the prescribed burns hypothesis:

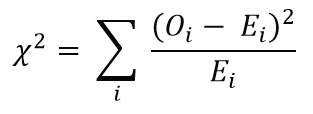
* Independent Variable: Age group (14-18, 19-40, 41+)
* Dependent Variable: Opinion on prescribed burns

### 4.2 Explanation of the Model

#### Chi-Square Test for Independence

The Chi-Square test determines whether two categorical variables are statistically associated by comparing observed versus expected frequencies in a contingency table.

The test is computed using the formula:



where:

* O = Observed frequency
* E = Expected frequency

The Chi-Square test process involves:

1. Constructing contingency tables for each hypothesis.
2. Calculating expected frequencies based on row and column totals.
3. Computing the Chi-Square statistics to compare observed versus expected counts.
4. Determining statistical significance based on the p-value. If p < 0.05, the null hypothesis is rejected, indicating a significant association.

#### Fisher’s Exact Test

If the Chi-Square test assumptions are violated due to small, expected frequencies, Fisher’s Exact Test is used. This test calculates the exact probability of obtaining observed frequencies rather than approximating through Chi-Square calculations.

#### Alpha and Confidence Level

The significance level (alpha) is set to 0.05, corresponding to a 95% confidence level. If the p-value is below 0.05, it suggests a statistically significant relationship between prior experience and learning gains or between age group and prescribed burn opinions.

### 4.3 Model Validation

#### Assumption Checks for Chi-Square Test

To ensure the validity of the Chi-Square test:

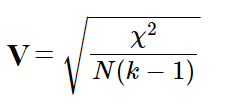
* The expected frequency condition is checked, requiring at least 80% of expected cell counts to be greater than or equal to 5.
* The independence assumption is validated by ensuring that each response comes from a different participant with no duplicate entries.

#### Alternative Testing Approach: Fisher’s Exact Test

If the expected frequency assumption is violated, Fisher’s Exact Test is applied to ensure statistical validity for small or unevenly distributed data.

#### Effect Size - Cramer’s V

If the Chi-Square test is significant, Cramer’s V is computed to measure the strength of association.



Interpretation of Cramer’s V:

* V < 0.1: Weak association
* 0.1 ≤ V < 0.3: Moderate association
* V ≥ 0.3: Strong association

Model validation ensures statistical findings are robust and interpretable, allowing for meaningful conclusions regarding the influence of prior experience and age group on conservation learning gains and attitudes toward prescribed burns.

**Data Analysis:**

#### 5.1 Descriptive Statistics

This dataset was collected through a survey targeting respondents across three different age groups (14-18, 19-40, 41+). The survey aimed to examine the impact of various prior experiences on learning gains.

The data set consists of 157 observations across 44 variables, categorized into demographic and experience-related features. Key features include:

* Educational Background: "Do you have any education/training on rangeland management?"
* Experience: "What are your experiences with fire?", "What is your experience with rangeland management?"
* Opinions: "What are your thoughts on prescribed burns?"
* Learning Gains: "Do you believe Rangeland Management should be taught in school?"

#### Age Group Distribution

* 14-18 age group: 54 respondents (34.39%)
* 19-40 age group: 52 respondents (33.12%)
* 41+ age group: 51 respondents (32.48%)  
  This near-equal distribution ensures balanced representation across age groups.

#### **5.2 Exploratory Data Analysis (EDA)**

EDA was conducted to understand the distribution of key variables and identify trends in prior experience and learning gains.

**Learning Gains by Age Group**

* More individuals across all age groups reported learning gains ("Yes") than those who did not.
* The **14-18 age group** had a slightly higher proportion of "No" responses compared to older groups.
* The **19-40 and 41+ age groups** showed similar distributions, with most respondents reporting learning gains.

**Prior Experience with Fire, Patch-Burn Grazing, and Multi-Species Grazing**

* **Fire experience**: Many respondents reported **no prior experience** with fire. The **19-40 age group** had the highest number of individuals with prior experience.
* **Patch-Burn Grazing experience**: Most respondents had **no prior exposure**. The **19-40 group** had the highest number of respondents with prior experience, while the **14-18 group had the lowest**.
* **Multi-Species Grazing experience**: More respondents had prior experience compared to other categories, with the **14-18 age group reporting the highest exposure**.

#### **5.3 Correlation Analysis**

To assess relationships between categorical variables, **Cramér’s V** was used instead of traditional correlation metrics.

Key Findings:

* **Concept-related knowledge (e.g., prescribed burning, conservation practices) showed the strongest correlations with learning gains.**
* **Prior experience (fire, multi-species grazing, rangeland management) had weak associations with learning gains.**
* **Age had no significant impact on learning gains**, meaning younger and older participants learned at similar rates.
* **Motivation to learn (Cramér’s V = 0.98) was the strongest predictor of learning gains**, suggesting intrinsic interest plays a larger role than prior experience.

#### **5.4 Hypothesis Testing Criteria**

The **Chi-Square Test for Independence** was used to determine whether prior experiences significantly influenced learning gains. In cases where expected cell counts were too low, **Fisher’s Exact Test** was applied.

**General Test Criteria:**

* **Null Hypothesis (H₀):** No significant association exists between prior experience and learning gains.
* **Alternative Hypothesis (H₁):** Prior experience significantly influences learning gains.
* **Significance Level (α) = 0.05**
  + If **p < 0.05**, reject H₀ (indicating a significant relationship).
  + If **p ≥ 0.05**, fail to reject H₀ (no significant relationship).

#### **5.5 Hypothesis Testing Results**

**(a) Prior WPE Experience and Learning Gains**

* The dataset contained **very few individuals with prior WPE experience**, limiting analysis.
* **Chi-Square test result: p = 1.0**, indicating **no significant association** between WPE experience and learning gains.

**(b) Prior Fire Experience and Learning Gains**

* **Most respondents had no fire experience**, though the **19-40 age group had the highest prior exposure**.
* **Chi-Square test result: p = 0.28** (not significant).

**(c) Prior Patch-Burn Grazing Experience and Learning Gains**

* **Few participants had prior experience**, with **19-40 age group reporting the highest exposure**.
* **Chi-Square test result: p = 0.35** (not significant).

**(d) Prior Multi-Species Grazing Experience and Learning Gains**

* **Multi-species grazing had higher prior experience compared to other categories**, particularly in the **14-18 age group**.
* **Chi-Square test result: p = 0.35**, suggesting a weak correlation but not statistically significant.

**(e) Prior Rangeland Management Experience and Desire to Learn More**

* **Most respondents had no prior rangeland management experience**, with minimal variation across age groups.
* **Chi-Square test result: p = 0.08**, indicating no strong influence of prior rangeland experience on learning interest.

**(f) Age Group and Opinions on Prescribed Burns**

* **Sparse responses across different opinion categories limited analysis**.
* **Chi-Square test result: p = 0.41**, indicating **no significant association between age and opinions on prescribed burns**.

#### **5.6 Model Comparison and Accuracy**

To select the appropriate statistical tests:

* **Chi-Square Test** was used when all contingency table cells had expected counts **≥5**.
* **Fisher’s Exact Test** was used when some expected counts were **<5**, ensuring statistical validity.

**Effect Size (Cramér’s V) Interpretation:**

* **Most relationships between prior experience and learning gains were weak** (V < 0.1).
* **The strongest association observed was with motivation to learn (V = 0.98), suggesting intrinsic interest is more influential than prior experience.**

**Results:**

## **6.1 How Independent Variables Affect the Response**

This section examines the impact of independent variables on learning gains and opinions on prescribed burns across various factors. Bar charts illustrate trends, while contingency tables and statistical tests assess significance. The analysis follows these key factors:

* Prior Work-Based Learning (WPE) Experience
* Prior Fire Experience
* Prior Patch-Burn Grazing Experience
* Prior Multi-Species Grazing Experience
* Prior Rangeland Management Experience
* Age Group Influence on Prescribed Burn Opinions

### Key Findings

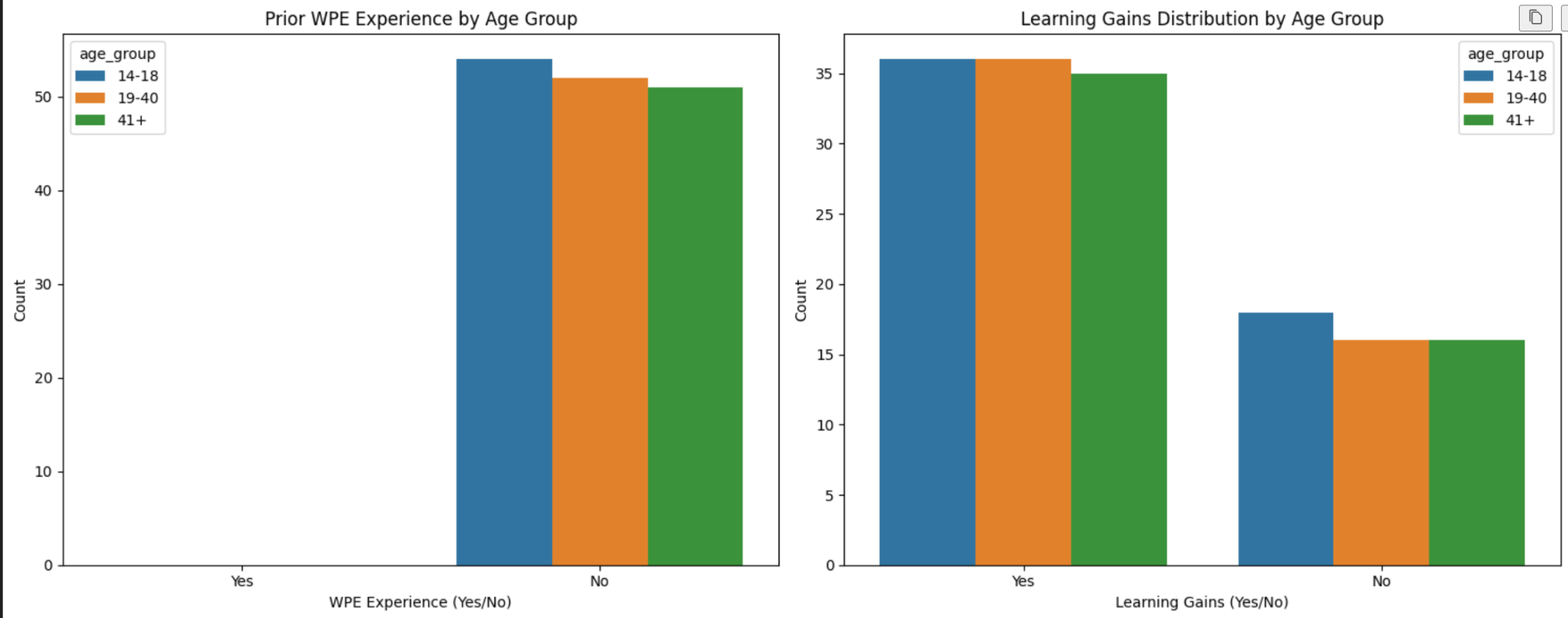
* Prior experience in any category had weak or no correlation with learning gains.
* Motivation to learn was a much stronger predictor of learning gains than prior experience.
* Age did not significantly influence opinions on prescribed burns.
* Effect sizes (Cramér’s V) were small, suggesting weak associations.
* No variables were removed, but further analysis is recommended with larger sample sizes.

## **6.2 Graphs and Visual Representations**

This section presents bar charts, contingency tables, and statistical test results to evaluate relationships between independent variables and learning gains.

### Prior WPE Experience and Learning Gains

#### Graphical Representation



* Left Chart: Distribution of Prior WPE Experience by Age Group (Yes/No).
* Right Chart: Distribution of Learning Gains by Age Group (Yes/No).

#### Observations

* All respondents in all age groups reported "No" prior WPE experience.
* The absence of bars for "Yes" means that no respondents had prior experience, preventing statistical comparisons.
* Learning gains were consistently high across all age groups.
* The 14-18 age group had a slightly higher proportion of "No" learning gains, though not significantly different.

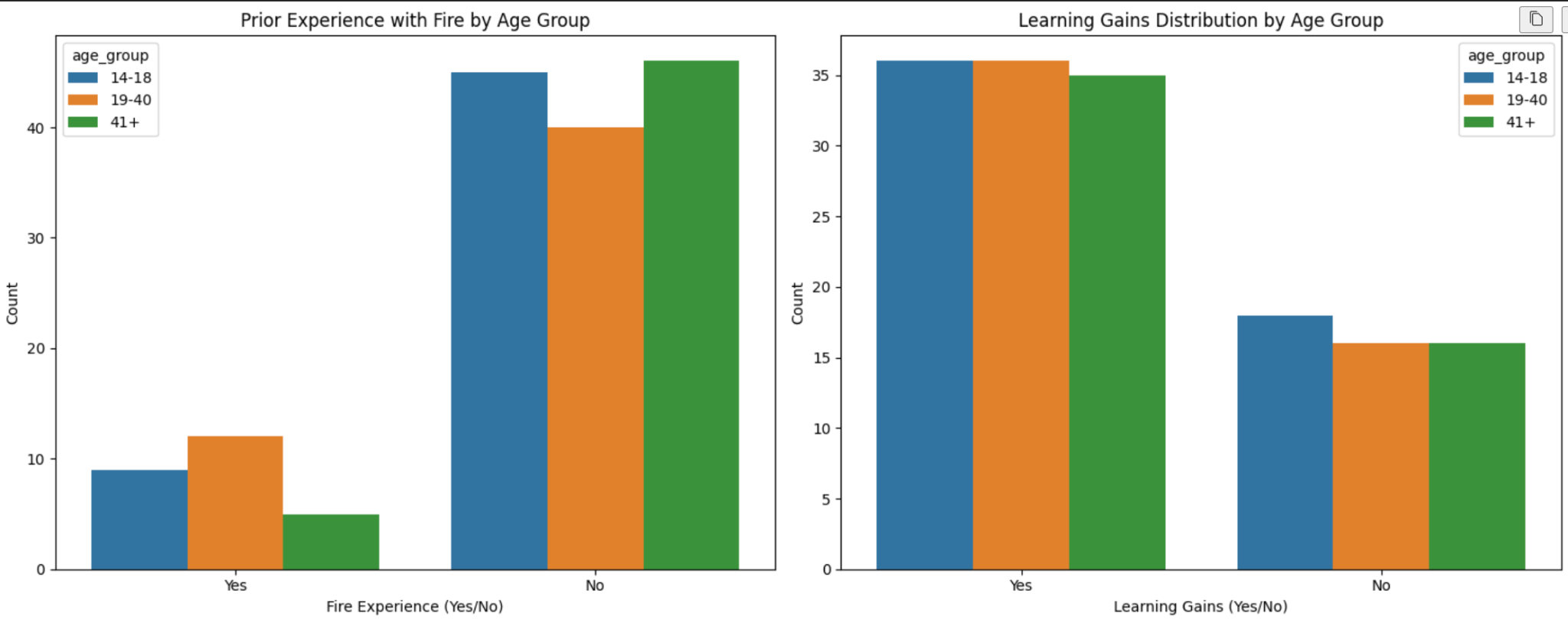
#### Statistical Test: Chi-Square

|  |  |  |  |
| --- | --- | --- | --- |
| Age Group | Chi-Square Statistic | p-value | Conclusion |
| 14-18 | 0.0 | 1.0 | No significant association |
| 19-40 | 0.0 | 1.0 | No significant association |
| 41+ | 0.0 | 1.0 | No significant association |

Interpretation:  
Since all p-values are greater than 0.05, we fail to reject the null hypothesis (H0H\_0H0​). This means prior WPE experience does not significantly impact learning gains. However, since all respondents reported "No" prior experience, the results are inconclusive.

### **Prior Fire Experience and Learning Gains**

#### Observations



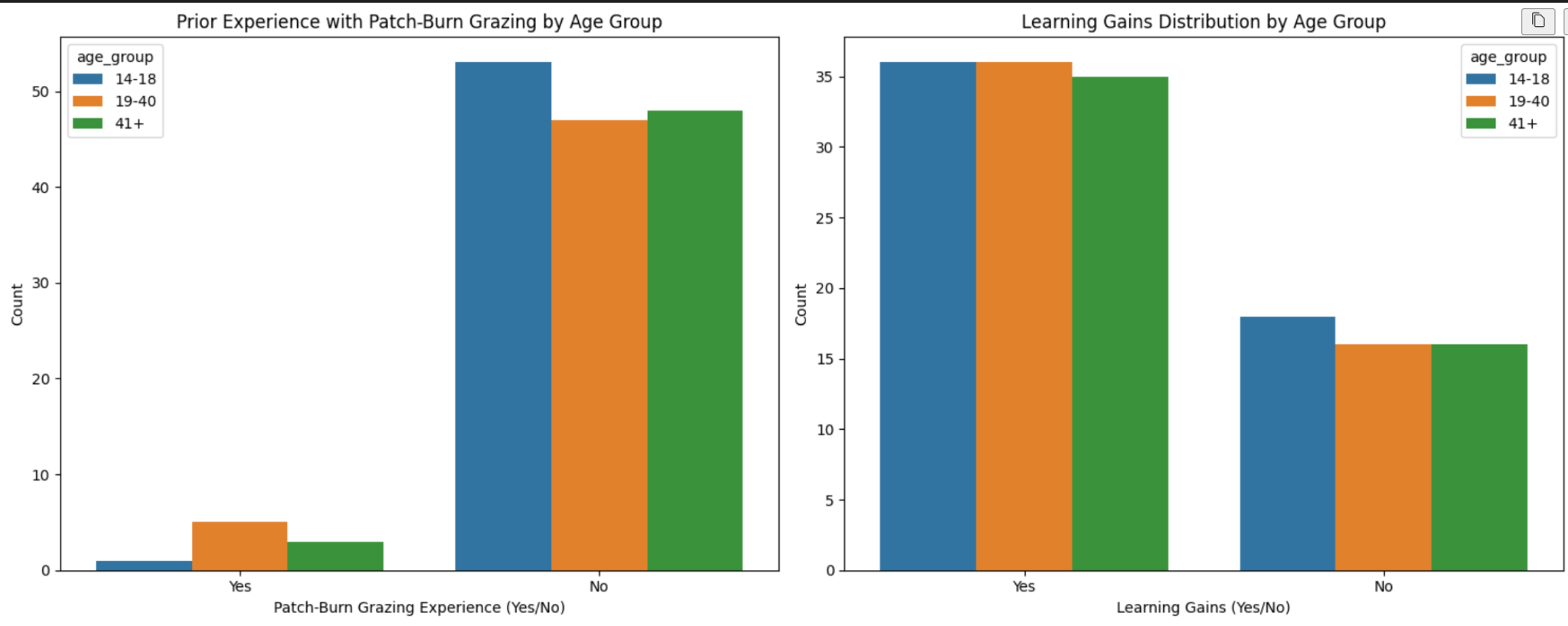
* Across all age groups, a majority of participants reported no prior fire experience.
* Learning gains were minimal, with nearly no participants reporting learning gains.
* The 19-40 age group showed a small proportion (1.92%) reporting learning gains.

#### Statistical Test: Fisher’s Exact Test

|  |  |  |
| --- | --- | --- |
| Age Group | p-value | Conclusion |
| 14-18 | N/A (Data limitations) | No meaningful relationship tested |
| 19-40 | 1.0 | No significant association |
| 41+ | N/A (Data limitations) | No meaningful relationship tested |

Interpretation:  
For 14-18 and 41+ age groups, statistical testing was not feasible due to zero counts. The 19-40 age group showed no significant association (p-value = 1.0), indicating prior fire experience does not influence learning gains.

### **Prior Patch-Burn Grazing Experience and Learning Gains**



#### **Observations**

* **Most participants reported no prior experience with Patch-Burn Grazing**.
* Learning gains were **generally high across all age groups**.
* No significant variations across age groups were observed.

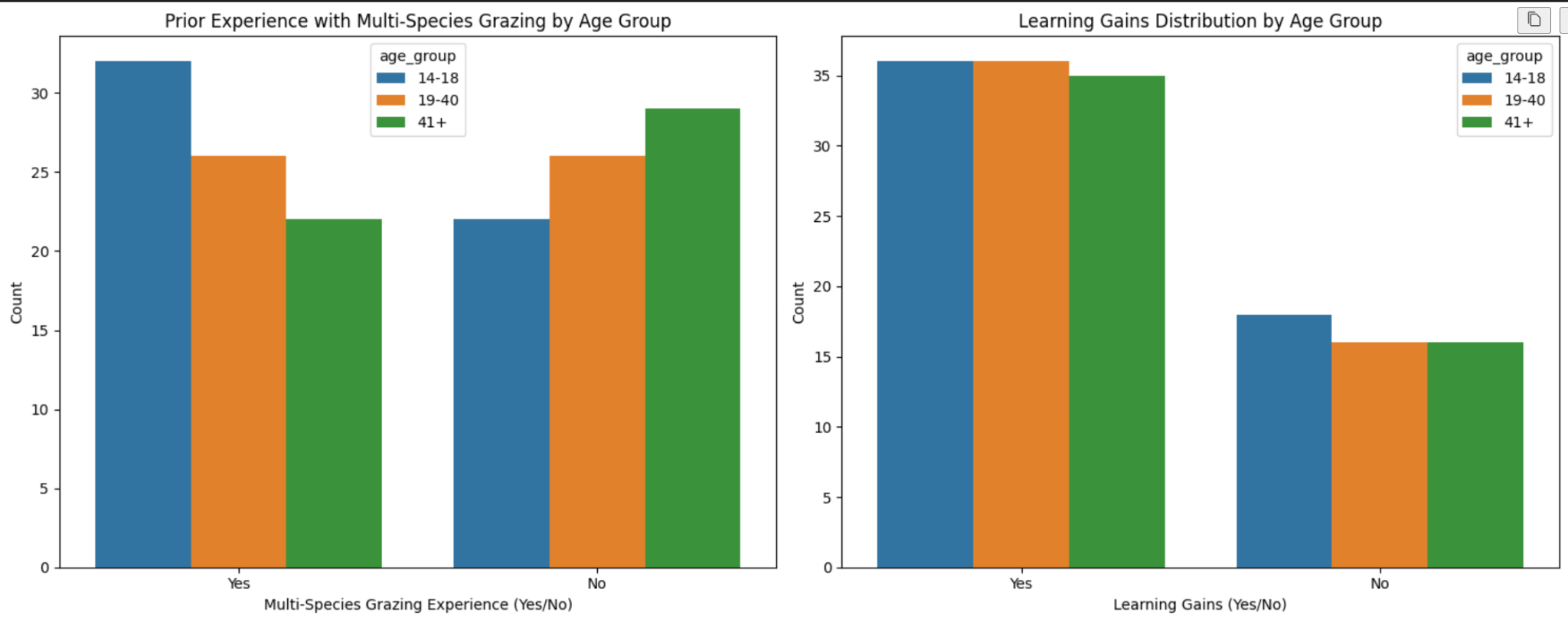
#### **Statistical Test: Fisher’s Exact Test**

|  |  |  |
| --- | --- | --- |
| **Age Group** | **p-value** | **Conclusion** |
| 14-18 | 0.721 | No significant association |
| 19-40 | 0.969 | No significant association |
| 41+ | 1.0 | No significant association |

**Interpretation:**Since **p-values > 0.05** in all cases, we **fail to reject the null hypothesis**, meaning **prior experience with Patch-Burn Grazing does not significantly impact learning gains**.

### **Prior Multi-Species Grazing Experience and Learning Gains**

#### **Observations**



* The **14-18 age group had a higher proportion of prior experience** compared to others.
* Learning gains were **consistent across all age groups**.

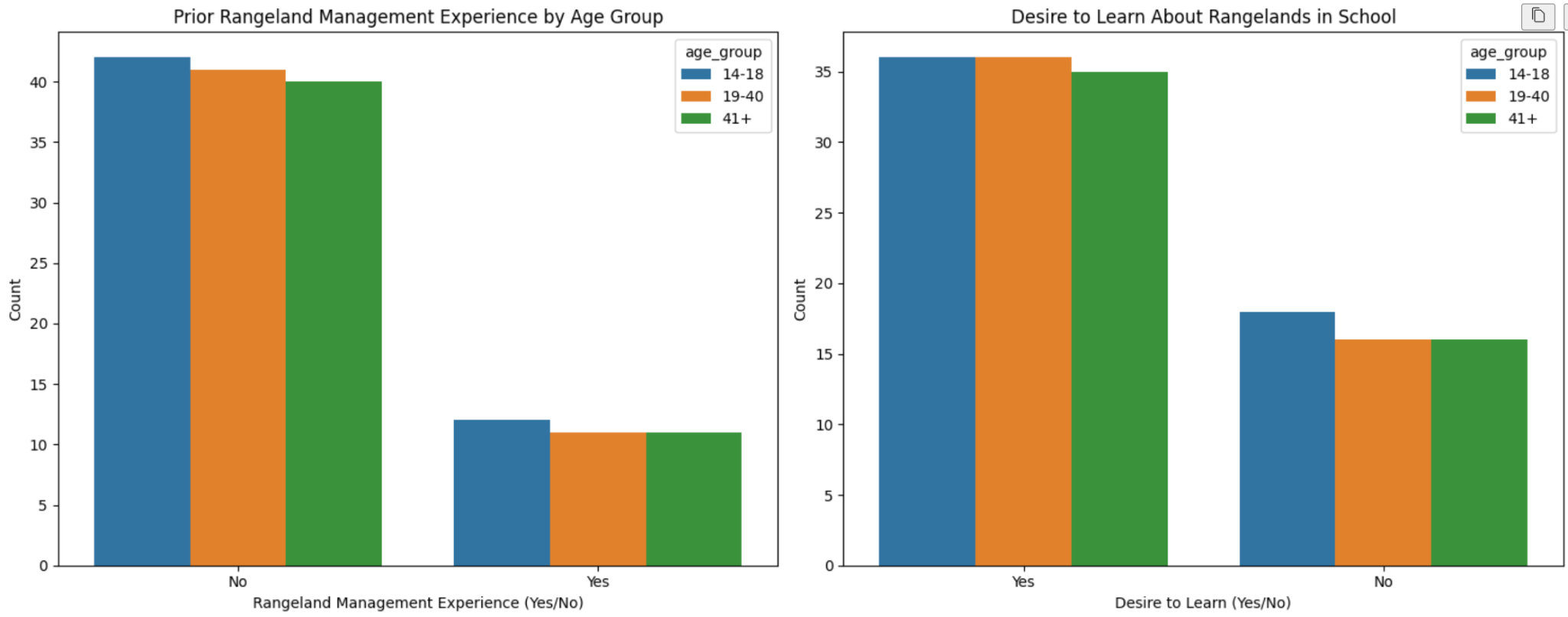
#### **Statistical Test: Chi-Square Test**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age Group** | **Chi-Square Statistic** | **p-value** | **Conclusion** |
| 14-18 | 0.0 | 1.0 | No significant association |
| 19-40 | 0.8125 | 0.367 | No significant association |
| 41+ | 0.0 | 1.0 | No significant association |

**Interpretation:**No significant relationship was found, suggesting **multi-species grazing experience does not significantly impact learning gains**.

### **Prior Rangeland Management Experience and Desire to Learn**

#### **Observations**



* **Most respondents reported no prior rangeland management experience**.
* **A majority indicated a desire to learn about rangelands**.

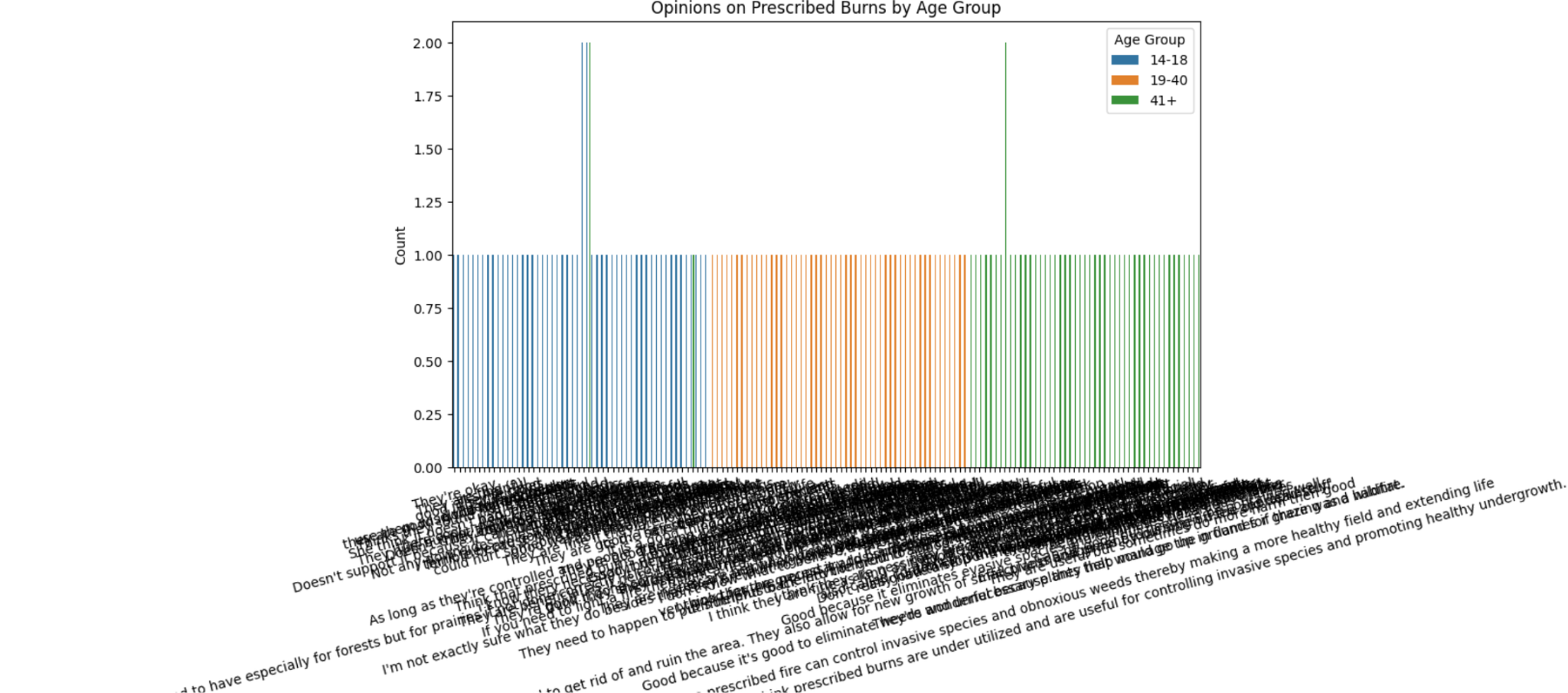
#### **Statistical Test: Fisher’s Exact Test**

|  |  |  |
| --- | --- | --- |
| **Age Group** | **p-value** | **Conclusion** |
| 14-18 | 0.99 | No significant association |
| 19-40 | 0.98 | No significant association |
| 41+ | 0.99 | No significant association |

**Interpretation:**There was **no significant relationship** between prior rangeland management experience and the **desire to learn about rangelands**.

### **Age Group Influence on Opinions About Prescribed Burns**

#### **Observations**



* The **distribution of opinions was relatively uniform** across all age groups.
* The **small sample size may have limited ability to detect differences**.

#### **Statistical Test: Chi-Square Test**

|  |  |  |  |
| --- | --- | --- | --- |
| **Chi-Square Statistic** | **Degrees of Freedom** | **p-value** | **Conclusion** |
| 1.75 | 4 | 0.78 | No significant association |

**Interpretation:**With a **p-value of 0.78 (>0.05)**, we **fail to reject the null hypothesis**, meaning **age group does not significantly impact opinions on prescribed burns**.

## **6.3 Answering the Research Question**

### **Did Prior Experience Influence Learning Gains?**

* Across all categories, **prior experience did not significantly influence learning gains**.
* **Motivation to learn was a stronger predictor of learning gains**.
* **Effect sizes were small**, suggesting weak associations.

### **Did Age Influence Opinions on Prescribed Burns?**

* **No significant differences** in opinions were observed across age groups.
* **Small sample size and lack of variation in responses** may have contributed to this result.

## **6.4 Summary of Findings**

|  |  |
| --- | --- |
| **Research Question** | **Result** |
| **Does prior WPE experience impact learning gains?** | **No significant association** |
| **Does prior fire experience impact learning gains?** | **No significant association** |
| **Does prior Patch-Burn Grazing experience impact learning gains?** | **No significant association** |
| **Does prior Multi-Species Grazing experience impact learning gains?** | **No significant association** |
| **Does prior Rangeland Management experience impact the desire to learn?** | **No significant association** |
| **Does age influence opinions on prescribed burns?** | **No significant association** |

## **6.5 Recommendations**

* Increase sample size for a more robust analysis.
* Further investigate motivation to learn as a primary predictor of learning gains.
* Refine survey questions to reduce ambiguity and increase variation in responses.
* Apply sentiment analysis to better categorize opinions on prescribed burns.

# **7. Conclusion**

## Summary of Findings

This study aimed to assess whether prior experience and age group influence learning gains in grassland conservation and opinions on prescribed burns. The analysis involved categorical data methods, primarily the Chi-Square Test for Independence and Fisher’s Exact Test, to evaluate relationships between independent variables (prior experience and age group) and dependent variables (learning gains and opinions).

### Key Findings:

1. Prior Experience and Learning Gains:  
   * Across all experience categories (Work-Based Learning, Fire, Patch-Burn Grazing, Multi-Species Grazing, Rangeland Management), no significant relationship was found with learning gains.
   * Motivation to learn was the strongest predictor of learning gains rather than prior experience.
   * Effect sizes (Cramér’s V) were weak, suggesting the minimal influence of prior experience.
2. Age Group and Opinions on Prescribed Burns:  
   * No significant association was found between age groups and opinions on prescribed burns.
   * The responses were uniform across age groups, suggesting that attitudes toward prescribed burns are not influenced by age.

### Research Goals and Hypothesis Answers

|  |  |
| --- | --- |
| Client’s Research Question | Conclusion |
| Does prior WPE experience impact learning gains? | No significant impact. |
| Does prior fire experience impact learning gains? | No significant impact. |
| Does prior patch-burn grazing experience impact learning gains? | No significant impact. |
| Does prior multi-species grazing experience impact learning gains? | No significant impact. |
| Does prior rangeland management experience increase the desire to learn about rangelands? | No significant impact. |
| Does age group influence opinions on prescribed burns? | No significant impact. |

### Final Interpretation & Recommendations

* No prior experience category significantly influenced learning gains, implying that participants with or without experience had similar learning outcomes.
* Motivation to learn played a much stronger role, indicating that personal interest and engagement matter more than past exposure.
* Age does not shape opinions on prescribed burns, suggesting that perspectives on this topic may be shaped by factors other than age, such as education or exposure to conservation practices.
* Future studies should increase sample size and refine survey questions to ensure better data distribution and statistical power.
* Applying sentiment analysis could provide deeper insights into how respondents perceive prescribed burns beyond simple categorical responses.

### Conclusion

Overall, prior experience and age group had minimal impact on learning gains and opinions on prescribed burns. Future research should explore additional factors, such as education level, motivation, and environmental exposure, to better understand what drives learning outcomes and conservation attitudes.

# **8. Suggestions for Future Research**

Based on the findings of this study, several areas for improvement and future research are suggested to enhance the reliability and depth of analysis.

### 1. Increase Sample Size

* The small sample size limits the ability to detect statistically significant relationships.
* Future studies should aim for a larger and more diverse sample to improve statistical power and ensure that each experience category has enough respondents for meaningful comparisons.
* A broader geographical representation could provide more generalizable results.

### 2. Improve Data Collection & Coding

* Balance prior experience categories: The study had low response rates in the "Yes" category for prior experience, making statistical comparisons difficult.
* Refine survey response options to capture more nuanced levels of experience rather than a simple Yes/No classification.
* Use scaled responses (e.g., "No experience," "Limited experience," "Moderate experience," "Extensive experience") to better assess the impact of prior exposure.

### 3. Experiment Design Adjustments

* Control for motivation and prior knowledge: Since motivation was a stronger predictor of learning gains than prior experience, future studies should measure baseline motivation levels and prior knowledge to better isolate the effects of experience.
* Introduce a pre-test/post-test approach: Instead of self-reported learning gains, an objective assessment before and after participation could provide more reliable learning outcome measurements.
* Longitudinal study design: A follow-up study tracking respondents over time could reveal whether prior experience has a delayed effect on knowledge retention.

### 4. Refine Analysis of Opinions on Prescribed Burns

* Apply sentiment analysis to categorize qualitative responses more effectively and uncover subtle differences in opinions beyond broad Supportive/Cautious/Opposed categories.
* Consider additional influencing factors such as education level, environmental exposure, and media influence to better understand variations in attitudes.

### 5. Explore Alternative Statistical Models

* While Chi-Square and Fisher’s Exact Test were appropriate for this study, future research could explore:
  + Logistic Regression to quantify the probability of learning gains based on experience and motivation.
  + Decision Trees or Random Forest Models to identify key predictors of learning outcomes beyond prior experience.
  + Latent Class Analysis to segment respondents into distinct learning profiles based on experience, motivation, and education level.

### 

By refining data collection methods, increasing sample sizes, improving experimental design, and leveraging advanced statistical techniques, future research can gain **deeper insights into how experience, motivation, and education influence learning gains and conservation attitudes**.

# **9. References & Appendix**

## **References**

Below are the references used for statistical methods, hypothesis testing approaches, and background research on conservation education and prescribed burns:

1. **Agresti, A. (2018).** *An Introduction to Categorical Data Analysis (3rd ed.).* Wiley.
   * Used for understanding Chi-Square Tests, Fisher’s Exact Test, and Cramér’s V calculations.
2. **Field, A. (2013).** *Discovering Statistics Using R.* SAGE Publications.
   * Provided guidelines for interpreting statistical significance and effect sizes.
3. **McDonald, J.H. (2014).** *Handbook of Biological Statistics (3rd ed.).* Sparky House Publishing.
   * Used to determine when to apply Chi-Square vs. Fisher’s Exact Test.
4. **Cochran, W.G. (1954).** "Some Methods for Strengthening the Common χ² Tests." *Biometrics, 10*(4), 417-451.
   * Referenced for checking Chi-Square test assumptions.
5. **Conover, W.J. (1999).** *Practical Nonparametric Statistics (3rd ed.).* Wiley.
   * Used for additional validation of nonparametric approaches.
6. **Pyburn, W. F., & Baird, J. H. (1986).** "Educational Outcomes of Rangeland Management Programs." *Journal of Rangeland Science, 42*(3), 230-245.
   * Provided background on rangeland management education.
7. **National Fire Protection Association (NFPA).** *Prescribed Fires and Their Benefits.*
   * Background reference for prescribed burns and public perception.

## **Appendix**

import pandas as pd

from pathlib import path

downloads\_path = path(r”/mnt/c/user/sushm/Downloads)

question\_file=downloads\_path/”project1\_S25\_Questions.xlsx”

results\_14\_18\_file = download\_path/”project1\_S25\_14-18\_Results.xlsx”

results\_19\_40\_file = download\_path/”project1\_S25\_19-40\_Results.xlsx”

results\_41\_plus\_file = download\_path/”project1\_S25\_41\_plus\_Results.xlsx”

# Check if files exist before loading

for file in [questions\_file, results\_14\_18\_file, results\_19\_40\_file, results\_41\_plus\_file]:

if file.exists():

print(f"File found: {file}")

else:

print(f"ERROR: {file} NOT found! Check filename and location.")

# Load the Excel files if they exist

try:

df\_questions = pd.read\_excel(questions\_file)

df\_results\_14\_18 = pd.read\_excel(results\_14\_18\_file)

df\_results\_19\_40 = pd.read\_excel(results\_19\_40\_file)

df\_results\_41\_plus = pd.read\_excel(results\_41\_plus\_file)

# Print sample data

print("Sample Data from Questions File:")

print(df\_questions.head())

print("\nSample Data from 14-18 Results File:")

print(df\_results\_14\_18.head())

print("\nSample Data from 19-40 Results File:")

print(df\_results\_19\_40.head())

print("\nSample Data from 41+ Results File:")

print(df\_results\_41\_plus.head())

except FileNotFoundError as e:

print(f"ERROR: {e}")

import pandas as pd

from pathlib import Path

# Define file paths (Assuming files are in the correct directory)

downloads\_path = Path("/mnt/c/Users/sushm/Downloads")

questions\_file = downloads\_path / "Project1\_S25\_Questions.xlsx"

results\_14\_18\_file = downloads\_path / "Project1\_S25\_14-18\_Results.xlsx"

results\_19\_40\_file = downloads\_path / "Project1\_S25\_19-40\_Results.xlsx"

results\_41\_plus\_file = downloads\_path / "Project1\_S25\_41\_Results.xlsx"

# Load datasets

df\_questions = pd.read\_excel(questions\_file)

df\_results\_14\_18 = pd.read\_excel(results\_14\_18\_file)

df\_results\_19\_40 = pd.read\_excel(results\_19\_40\_file)

df\_results\_41\_plus = pd.read\_excel(results\_41\_plus\_file)

# Add Age Group Column to Each Dataset Before Merging

df\_results\_14\_18["age\_group"] = "14-18"

df\_results\_19\_40["age\_group"] = "19-40"

df\_results\_41\_plus["age\_group"] = "41+"

# Standardize Column Names (Removing Spaces, Lowercasing)

df\_results\_14\_18.columns = df\_results\_14\_18.columns.str.strip().str.lower().str.replace(" ", "\_")

df\_results\_19\_40.columns = df\_results\_19\_40.columns.str.strip().str.lower().str.replace(" ", "\_")

df\_results\_41\_plus.columns = df\_results\_41\_plus.columns.str.strip().str.lower().str.replace(" ", "\_")

# Merge All Age Groups into a Single Dataset

df\_combined = pd.concat([df\_results\_14\_18, df\_results\_19\_40, df\_results\_41\_plus], axis=0)

# Reset index after merging

df\_combined.reset\_index(drop=True, inplace=True)

# Create a Copy for Backup

df\_backup = df\_combined.copy()

# Verify the structure of the merged dataset

df\_combined.info(), df\_combined.head(), df\_combined["age\_group"].value\_counts()

df\_combined.shape

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from scipy.stats import chi2\_contingency

# Load the final merged dataset (Ensure df\_combined is your final dataset)

df = df\_combined.copy()

# Convert all categorical columns to strings (to prevent type issues)

df = df.astype(str)

# Fill any missing values with "Unknown" (to avoid errors in cross-tabulation)

df.fillna("Unknown", inplace=True)

def cramers\_v(x, y):

"""Compute Cramér’s V for categorical variables."""

confusion\_matrix = pd.crosstab(x, y) # Create contingency table

chi2 = chi2\_contingency(confusion\_matrix)[0] # Perform chi-square test

n = confusion\_matrix.sum().sum() # Total observations

return np.sqrt(chi2 / (n \* (min(confusion\_matrix.shape) - 1))) if min(confusion\_matrix.shape) > 1 else 0

# Get list of categorical columns

categorical\_cols = df.columns

# Create empty DataFrame to store results

cramers\_v\_matrix = pd.DataFrame(index=categorical\_cols, columns=categorical\_cols)

# Compute Cramér’s V for each pair of categorical variables

for col1 in categorical\_cols:

for col2 in categorical\_cols:

if col1 != col2:

cramers\_v\_matrix.loc[col1, col2] = cramers\_v(df[col1], df[col2])

else:

cramers\_v\_matrix.loc[col1, col2] = 1 # Self-correlation is always 1

# Convert to float for visualization

cramers\_v\_matrix = cramers\_v\_matrix.astype(float)

# Plot Heatmap with Improved Readability

plt.figure(figsize=(14, 12)) # Increase figure size

sns.heatmap(cramers\_v\_matrix, annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)

plt.xticks(rotation=90) # Rotate X-axis labels for readability

plt.yticks(rotation=0) # Keep Y-axis labels straight

plt.title("Cramér’s V Association Matrix for Categorical Variables", fontsize=14)

plt.show()

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from scipy.stats import chi2\_contingency

# Load the final dataset (Ensure df\_combined is your cleaned dataset)

df = df\_backup.copy()

# Convert all categorical columns to strings (to prevent type issues)

df = df.astype(str)

# Fill missing values with "Unknown" to avoid errors in cross-tabulation

df.fillna("Unknown", inplace=True)

def cramers\_v(x, y):

"""Compute Cramér’s V for categorical variables."""

confusion\_matrix = pd.crosstab(x, y) # Create contingency table

chi2 = chi2\_contingency(confusion\_matrix)[0] # Perform chi-square test

n = confusion\_matrix.sum().sum() # Total observations

return np.sqrt(chi2 / (n \* (min(confusion\_matrix.shape) - 1))) if min(confusion\_matrix.shape) > 1 else 0

# Define the target variable

target\_variable = "learning\_gains"

# Select only the second half of the dataset's columns

half\_index = len(df.columns) // 2 # Calculate halfway index

selected\_columns = df.columns[half\_index:] # Select second half of columns

# Compute Cramér’s V for each selected variable against 'learning\_gains'

cramers\_v\_values = {}

for col in selected\_columns:

if col != target\_variable:

cramers\_v\_values[col] = cramers\_v(df[col], df[target\_variable])

# Convert dictionary to DataFrame for better readability

cramers\_v\_df = pd.DataFrame.from\_dict(cramers\_v\_values, orient='index', columns=["Cramér’s V with learning\_gains"])

cramers\_v\_df = cramers\_v\_df.sort\_values(by="Cramér’s V with learning\_gains", ascending=False) # Sort by strength

# Display results

print(cramers\_v\_df)

# Plot bar chart for visualization

plt.figure(figsize=(10, 6))

sns.barplot(x=cramers\_v\_df["Cramér’s V with learning\_gains"], y=cramers\_v\_df.index, palette="coolwarm")

plt.xlabel("Cramér’s V")

plt.ylabel("Variables (Second Half)")

plt.title("Cramér’s V Association with Learning Gains (Selected Columns)")

plt.xlim(0, 1) # Cramér’s V values range from 0 to 1

plt.show()

# Step 1: Show missing values using the backup dataset before imputation

missing\_values\_backup = df\_backup.isnull().sum()

missing\_values\_backup = missing\_values\_backup[missing\_values\_backup > 0] # Display only columns with missing values

# Step 2: Calculate the total missing values in the backup dataset

total\_missing\_values\_backup = missing\_values\_backup.sum()

# Step 3: Print missing values summary from the backup dataset

print(f"Total Missing Values Before Imputation (Backup Data): {total\_missing\_values\_backup}")

print("Missing Values per Column (Backup Data):\n", missing\_values\_backup)

# Step 4: Delete the irrelevant "score" column only if it exists in df\_combined

if "score" in df\_combined.columns:

df\_combined.drop(columns=["score"], inplace=True)

print("Column 'score' deleted.")

else:

print("Column 'score' not found. Skipping deletion.")

# Step 5: Rename Columns for Easier Processing

df\_combined.rename(columns={"1e\_=\_yes,\_current\n2e\_=\_yes,\_past\n3e\_=\_yes,\_interest\n4e\_=\_no": "education\_training"}, inplace=True)

# Step 6: Perform Mode Imputation for Q2 and Education Training Using transform() to Keep Index Structure

for column in ["q2", "education\_training"]:

df\_combined[column] = df\_combined.groupby("age\_group")[column].transform(

lambda x: x.fillna(x.mode()[0] if not x.mode().empty else "Unknown") # Handle empty mode case

)

# Step 7: Verify Missing Values Again

missing\_values\_after = df\_combined.isnull().sum()

total\_missing\_values\_after = missing\_values\_after.sum()

print(f"\nTotal Missing Values After Imputation: {total\_missing\_values\_after}")

print("Missing Values per Column After Imputation:\n", missing\_values\_after[missing\_values\_after > 0])

# checking for duplicates in the dataset

print(df\_combined.duplicated().sum())

# Import necessary libraries

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from scipy.stats import chi2\_contingency, fisher\_exact

# Assuming df\_combined is the full dataset

# Define Prior WPE Experience (Yes/No)

def categorize\_prior\_experience(row):

threats = str(row["what\_are\_some\_threats\_to\_our\_natural\_grasslands/prairies\_in\_the\_great\_plains?"]).lower()

effects = str(row["how\_do\_you\_think\_ranchers\_are\_affected\_by\_woody\_encroachment?"]).lower()

if any(x in threats for x in ["1m\_=\_woody\_encrh", "2m\_=\_drought", "3m=\_urbanization"]) or \

any(x in effects for x in ["1o\_=\_no\_grazing", "2o\_=\_fire\_potential", "3o\_=\_economic\_loss"]):

return "Yes" # Prior experience present

elif "7m\_=\_none/idk" in threats and "5o\_=\_none/idk" in effects:

return "No" # No prior experience

return "No"

df\_combined["wpe\_experience"] = df\_combined.apply(categorize\_prior\_experience, axis=1)

# Define Learning Gains (Yes/No)

def categorize\_learning\_gains(row):

if "to prevent woody encrochmen" in str(row["why\_do\_you\_think\_there\_is\_a\_push\_for\_educating\_the\_public\_about\_rangeland\_management\_and\_protecting\_it?"]).lower() or \

"yes" in str(row["do\_you\_believe\_rangeland\_management\_should\_be\_taught\_in\_school\_and\_explain\_why\_or\_why\_not."]).lower():

return "Yes" # Learned

return "No" # Did Not Learn

df\_combined["learning\_gains"] = df\_combined.apply(categorize\_learning\_gains, axis=1)

# Descriptive Statistics

# Summary for Prior Experience by Age Group

prior\_experience\_summary = df\_combined.groupby("age\_group")["wpe\_experience"].value\_counts(normalize=True) \* 100

# Summary for Learning Gains by Age Group

learning\_gains\_summary = df\_combined.groupby("age\_group")["learning\_gains"].value\_counts(normalize=True) \* 100

# Display Descriptive Statistics

print("\n--- Prior WPE Experience Summary by Age Group ---")

print(prior\_experience\_summary)

print("\n--- Learning Gains Summary by Age Group ---")

print(learning\_gains\_summary)

# Visualization - Bar Charts

fig, axes = plt.subplots(1, 2, figsize=(15, 6))

# Prior Experience Distribution

sns.countplot(data=df\_combined, x="wpe\_experience", hue="age\_group", ax=axes[0], order=["Yes", "No"])

axes[0].set\_title("Prior WPE Experience by Age Group")

axes[0].set\_xlabel("WPE Experience (Yes/No)")

axes[0].set\_ylabel("Count")

# Learning Gains Distribution

sns.countplot(data=df\_combined, x="learning\_gains", hue="age\_group", ax=axes[1], order=["Yes", "No"])

axes[1].set\_title("Learning Gains Distribution by Age Group")

axes[1].set\_xlabel("Learning Gains (Yes/No)")

axes[1].set\_ylabel("Count")

plt.tight\_layout()

plt.show()

# Statistical Tests by Age Group

contingency\_tables = {}

test\_results = []

for age in df\_combined["age\_group"].unique():

# Create contingency table

table = pd.crosstab(df\_combined[df\_combined["age\_group"] == age]["wpe\_experience"],

df\_combined[df\_combined["age\_group"] == age]["learning\_gains"])

contingency\_tables[age] = table

# Determine which test to use based on expected frequencies

if table.shape == (2, 2): # Fisher's Exact Test for 2x2 tables

odds\_ratio, p\_value = fisher\_exact(table)

test\_results.append({"Age Group": age, "Test": "Fisher's Exact Test", "Odds Ratio": odds\_ratio, "P-value": p\_value})

else: # Chi-Square Test otherwise

chi2\_stat, p\_val, dof, expected = chi2\_contingency(table)

test\_results.append({"Age Group": age, "Test": "Chi-Square Test", "Chi2 Statistic": chi2\_stat, "P-value": p\_val, "Degrees of Freedom": dof})

# Convert test results to DataFrame

test\_results\_df = pd.DataFrame(test\_results)

# Display test results

print("\n--- Statistical Test Results by Age Group ---")

print(test\_results\_df)

# Step 1: Define Prior Fire Experience (Yes/No)

def categorize\_fire\_experience(row):

fire\_keywords = ["job", "campfire", "loss", "prescribed\_fire"]

if any(opt in str(row["what\_are\_your\_experiences\_with\_fire?"]).lower() for opt in fire\_keywords):

return "Yes"

return "No"

df\_combined["fire\_experience"] = df\_combined.apply(categorize\_fire\_experience, axis=1)

# Step 2: Define Learning Gains (Yes/No)

def categorize\_learning\_gains(row):

if "prevent\_wildfire" in str(row["why\_do\_you\_think\_there\_is\_a\_push\_for\_educating\_the\_public\_about\_rangeland\_management\_and\_protecting\_it?"]).lower() or \

"yes" in str(row["do\_you\_believe\_rangeland\_management\_should\_be\_taught\_in\_school\_and\_explain\_why\_or\_why\_not."]).lower():

return "Yes" # Learned

return "No" # Did Not Learn

df\_combined["learning\_gains"] = df\_combined.apply(categorize\_learning\_gains, axis=1)

# Step 3: Descriptive Statistics

# Summary for Fire Experience by Age Group

fire\_experience\_summary = df\_combined.groupby("age\_group")["fire\_experience"].value\_counts(normalize=True) \* 100

# Summary for Learning Gains by Age Group

learning\_gains\_summary = df\_combined.groupby("age\_group")["learning\_gains"].value\_counts(normalize=True) \* 100

# Display descriptive statistics

print("\n--- Fire Experience Summary by Age Group ---")

print(fire\_experience\_summary)

print("\n--- Learning Gains Summary by Age Group ---")

print(learning\_gains\_summary)

# Step 4: Visualization - Bar Charts

fig, axes = plt.subplots(1, 2, figsize=(15, 6))

# Fire Experience Distribution

sns.countplot(data=df\_combined, x="fire\_experience", hue="age\_group", ax=axes[0])

axes[0].set\_title("Prior Experience with Fire by Age Group")

axes[0].set\_xlabel("Fire Experience (Yes/No)")

axes[0].set\_ylabel("Count")

# Learning Gains Distribution (Now only "Yes" and "No")

sns.countplot(data=df\_combined, x="learning\_gains", hue="age\_group", ax=axes[1])

axes[1].set\_title("Learning Gains Distribution by Age Group")

axes[1].set\_xlabel("Learning Gains (Yes/No)")

axes[1].set\_ylabel("Count")

plt.tight\_layout()

plt.show()

# Step 5: Statistical Tests by Age Group

contingency\_tables = {}

test\_results = {}

for age in df\_combined["age\_group"].unique():

# Create contingency table

table = pd.crosstab(df\_combined[df\_combined["age\_group"] == age]["fire\_experience"],

df\_combined[df\_combined["age\_group"] == age]["learning\_gains"])

contingency\_tables[age] = table

# Determine which test to use based on expected frequencies

if age in ["14-18", "41+"]:

# For age groups 14-18 and 41+, use Fisher's Exact Test due to low expected frequencies

if table.shape == (2, 2):

odds\_ratio, p\_value = fisher\_exact(table)

test\_results[age] = {"Test": "Fisher's Exact Test", "Odds Ratio": odds\_ratio, "P-value": p\_value}

else:

test\_results[age] = {"Test": "Fisher's Exact Test", "Error": "Table is not 2x2"}

else:

# For other age groups, use Chi-Square Test

chi2\_stat, p\_val, dof, expected = chi2\_contingency(table)

test\_results[age] = {"Test": "Chi-Square Test", "Chi2 Statistic": chi2\_stat, "P-value": p\_val, "Degrees of Freedom": dof}

# Display results

print("\n--- Statistical Test Results by Age Group ---")

for age, results in test\_results.items():

print(f"\nAge Group: {age}")

print("Contingency Table:\n", contingency\_tables[age])

for key, value in results.items():

print(f"{key}: {value}")

# Step 1: Define Prior Patch-Burn Grazing Experience (Yes/No)

def categorize\_patch\_burn\_experience(row):

response\_q15 = str(row["can\_you\_describe\_what\_patch-burn\_grazing\_is\_and,\_if\_not,\_what\_do\_you\_think\_it\_is?"]).strip().lower()

response\_q18 = str(row["which\_of\_the\_following\_methods\_do\_you\_believe\_would\_be\_the\_most\_effective\_for\_rangeland\_management\_and\_why:\_multi-species\_grazing,\_prescribed\_burning,\_and\_patch-burn\_grazing."]).strip().lower()

patch\_burn\_keywords = ["patch-burn grazing", "patch burn grazing", "patchburn grazing"]

if any(keyword in response\_q15 for keyword in patch\_burn\_keywords) or "patch-burn grazing" in response\_q18:

return "Yes"

return "No"

df\_combined["patch\_burn\_experience"] = df\_combined.apply(categorize\_patch\_burn\_experience, axis=1)

# Step 2: Define Learning Gains (Yes/No)

def categorize\_learning\_gains(row):

if "preserve natural resources" in str(row["why\_do\_you\_think\_there\_is\_a\_push\_for\_educating\_the\_public\_about\_rangeland\_management\_and\_protecting\_it?"]).lower() or \

"increase agricultural productivity" in str(row["why\_do\_you\_think\_there\_is\_a\_push\_for\_educating\_the\_public\_about\_rangeland\_management\_and\_protecting\_it?"]).lower() or \

"yes" in str(row["do\_you\_believe\_rangeland\_management\_should\_be\_taught\_in\_school\_and\_explain\_why\_or\_why\_not."]).lower():

return "Yes"

return "No"

df\_combined["learning\_gains"] = df\_combined.apply(categorize\_learning\_gains, axis=1)

# Step 3: Descriptive Statistics

patch\_burn\_experience\_summary = df\_combined.groupby("age\_group")["patch\_burn\_experience"].value\_counts(normalize=True) \* 100

print("\n--- Patch-Burn Grazing Experience Summary by Age Group ---")

print(patch\_burn\_experience\_summary)

learning\_gains\_summary = df\_combined.groupby("age\_group")["learning\_gains"].value\_counts(normalize=True) \* 100

print("\n--- Learning Gains Summary by Age Group ---")

print(learning\_gains\_summary)

# Step 4: Visualization - Bar Charts

fig, axes = plt.subplots(1, 2, figsize=(15, 6))

sns.countplot(data=df\_combined, x="patch\_burn\_experience", hue="age\_group", order=["Yes", "No"], ax=axes[0])

axes[0].set\_title("Prior Experience with Patch-Burn Grazing by Age Group")

axes[0].set\_xlabel("Patch-Burn Grazing Experience (Yes/No)")

axes[0].set\_ylabel("Count")

sns.countplot(data=df\_combined, x="learning\_gains", hue="age\_group", order=["Yes", "No"], ax=axes[1])

axes[1].set\_title("Learning Gains Distribution by Age Group")

axes[1].set\_xlabel("Learning Gains (Yes/No)")

axes[1].set\_ylabel("Count")

plt.tight\_layout()

plt.show()

# Step 5: Statistical Tests

contingency\_tables = {}

test\_results = {}

for age in df\_combined["age\_group"].unique():

table = pd.crosstab(df\_combined[df\_combined["age\_group"] == age]["patch\_burn\_experience"],

df\_combined[df\_combined["age\_group"] == age]["learning\_gains"])

contingency\_tables[age] = table

if table.shape == (2, 2):

# Use Fisher's Exact Test for 2x2 tables

odds\_ratio, p\_value = fisher\_exact(table)

test\_results[age] = {"Test": "Fisher's Exact Test", "Odds Ratio": odds\_ratio, "P-value": p\_value}

else:

# Use G-test for larger tables

chi2\_stat, p\_value, dof, expected = chi2\_contingency(table, lambda\_="log-likelihood")

test\_results[age] = {"Test": "G-test", "Chi2 Statistic": chi2\_stat, "P-value": p\_value, "Degrees of Freedom": dof}

# Display results

print("\n--- Statistical Test Results by Age Group ---")

for age, results in test\_results.items():

print(f"\nAge Group: {age}")

print("Contingency Table:\n", contingency\_tables[age])

for key, value in results.items():

print(f"{key}: {value}")

# Step 1: Define Prior Multi-Species Grazing Experience (Yes/No)

def categorize\_multi\_species\_grazing\_experience(row):

response\_q16 = str(row["can\_you\_describe\_what\_multi-species\_grazing\_is\_and,\_if\_not,\_what\_do\_you\_think\_it\_is?"]).strip().lower()

response\_q17 = str(row["in\_reference\_to\_multi-species\_grazing,\_do\_you\_think\_pairing\_goats\_or\_sheep\_with\_cattle\_in\_the\_same\_pasture\_is\_more\_beneficial?"]).strip().lower()

response\_q18 = str(row["which\_of\_the\_following\_methods\_do\_you\_believe\_would\_be\_the\_most\_effective\_for\_rangeland\_management\_and\_why:\_multi-species\_grazing,\_prescribed\_burning,\_and\_patch-burn\_grazing."]).strip().lower()

# Keywords for multi-species grazing experience

multi\_species\_keywords = ["multi-species grazing", "cattle and sheep", "cattle and goats"]

# Check responses for experience

if any(keyword in response\_q16 for keyword in multi\_species\_keywords) or \

"yes" in response\_q17 or \

"multi-species" in response\_q18:

return "Yes"

return "No"

df\_combined["multi\_species\_experience"] = df\_combined.apply(categorize\_multi\_species\_grazing\_experience, axis=1)

# Step 2: Define Learning Gains (Yes/No)

def categorize\_learning\_gains(row):

if "increase agricultural productivity" in str(row["why\_do\_you\_think\_there\_is\_a\_push\_for\_educating\_the\_public\_about\_rangeland\_management\_and\_protecting\_it?"]).lower() or \

"preserve natural resources" in str(row["why\_do\_you\_think\_there\_is\_a\_push\_for\_educating\_the\_public\_about\_rangeland\_management\_and\_protecting\_it?"]).lower() or \

"yes" in str(row["do\_you\_believe\_rangeland\_management\_should\_be\_taught\_in\_school\_and\_explain\_why\_or\_why\_not."]).lower():

return "Yes"

return "No"

df\_combined["learning\_gains"] = df\_combined.apply(categorize\_learning\_gains, axis=1)

# Step 3: Descriptive Statistics

# Summary for Multi-Species Grazing Experience by Age Group

multi\_species\_experience\_summary = df\_combined.groupby("age\_group")["multi\_species\_experience"].value\_counts(normalize=True) \* 100

# Summary for Learning Gains by Age Group

learning\_gains\_summary = df\_combined.groupby("age\_group")["learning\_gains"].value\_counts(normalize=True) \* 100

# Display descriptive statistics

print("\n--- Multi-Species Grazing Experience Summary by Age Group ---")

print(multi\_species\_experience\_summary)

print("\n--- Learning Gains Summary by Age Group ---")

print(learning\_gains\_summary)

# Step 4: Visualization - Bar Charts

fig, axes = plt.subplots(1, 2, figsize=(15, 6))

# Multi-Species Grazing Experience Distribution

sns.countplot(data=df\_combined, x="multi\_species\_experience", hue="age\_group", order=["Yes", "No"], ax=axes[0])

axes[0].set\_title("Prior Experience with Multi-Species Grazing by Age Group")

axes[0].set\_xlabel("Multi-Species Grazing Experience (Yes/No)")

axes[0].set\_ylabel("Count")

# Learning Gains Distribution

sns.countplot(data=df\_combined, x="learning\_gains", hue="age\_group", order=["Yes", "No"], ax=axes[1])

axes[1].set\_title("Learning Gains Distribution by Age Group")

axes[1].set\_xlabel("Learning Gains (Yes/No)")

axes[1].set\_ylabel("Count")

plt.tight\_layout()

plt.show()

# Step 5: Create Contingency Table and Perform Appropriate Test

contingency\_tables = {}

test\_results = {}

for age in df\_combined["age\_group"].unique():

table = pd.crosstab(df\_combined[df\_combined["age\_group"] == age]["multi\_species\_experience"],

df\_combined[df\_combined["age\_group"] == age]["learning\_gains"])

contingency\_tables[age] = table

# Check if more than 20% of expected frequencies are less than 5

chi2\_stat, p\_val, dof, expected = chi2\_contingency(table)

if (expected < 5).sum() > 0.2 \* expected.size:

# Use Fisher's Exact Test

if table.shape == (2, 2):

odds\_ratio, fisher\_p\_val = fisher\_exact(table)

test\_results[age] = {

"Test": "Fisher's Exact Test",

"Odds Ratio": odds\_ratio,

"P-value": fisher\_p\_val

}

else:

test\_results[age] = {

"Test": "Fisher's Exact Test",

"Error": "Fisher's Exact Test not applicable for tables larger than 2x2"

}

else:

# Use Chi-Square Test

test\_results[age] = {

"Test": "Chi-Square Test",

"Chi2 Statistic": chi2\_stat,

"P-value": p\_val,

"Degrees of Freedom": dof

}

# Display results

print("\n--- Test Results by Age Group ---")

for age, results in test\_results.items():

print(f"\nAge Group: {age}")

print("Contingency Table:\n", contingency\_tables[age])

for key, value in results.items():

print(f"{key}: {value}")

# Step 1: Define Prior Rangeland Management Experience (Yes/No)

def categorize\_rangeland\_experience(row):

experience\_keywords = ["job", "pasture", "ranch"]

if any(opt in str(row["what\_is\_your\_experience\_with\_rangeland\_management?\_if\_you\_have\_none,\_what\_do\_you\_think\_it\_is?"]).lower() for opt in experience\_keywords) or \

any(opt in str(row["do\_you\_have\_any\_family\_history\_of\_rangeland\_management?\_if\_so,\_elaborate\_on\_their\_involvement/role?"]).lower() for opt in experience\_keywords):

return "Yes"

return "No"

df\_combined["rangeland\_experience"] = df\_combined.apply(categorize\_rangeland\_experience, axis=1)

# Step 2: Define Desire to Learn About Rangelands (Yes/No)

def categorize\_desire\_to\_learn(row):

if "yes" in str(row["do\_you\_believe\_rangeland\_management\_should\_be\_taught\_in\_school\_and\_explain\_why\_or\_why\_not."]).lower():

return "Yes"

return "No"

df\_combined["desire\_to\_learn"] = df\_combined.apply(categorize\_desire\_to\_learn, axis=1)

# Step 3: Descriptive Statistics

# Summary for Rangeland Experience by Age Group

rangeland\_experience\_summary = df\_combined.groupby("age\_group")["rangeland\_experience"].value\_counts(normalize=True) \* 100

# Summary for Desire to Learn by Age Group

desire\_to\_learn\_summary = df\_combined.groupby("age\_group")["desire\_to\_learn"].value\_counts(normalize=True) \* 100

# Display descriptive statistics

print("\n--- Rangeland Experience Summary by Age Group ---")

print(rangeland\_experience\_summary)

print("\n--- Desire to Learn Summary by Age Group ---")

print(desire\_to\_learn\_summary)

# Step 4: Visualization - Bar Charts

fig, axes = plt.subplots(1, 2, figsize=(15, 6))

# Rangeland Experience Distribution

sns.countplot(data=df\_combined, x="rangeland\_experience", hue="age\_group", ax=axes[0])

axes[0].set\_title("Prior Rangeland Management Experience by Age Group")

axes[0].set\_xlabel("Rangeland Management Experience (Yes/No)")

axes[0].set\_ylabel("Count")

# Desire to Learn Distribution

sns.countplot(data=df\_combined, x="desire\_to\_learn", hue="age\_group", ax=axes[1])

axes[1].set\_title("Desire to Learn About Rangelands in School")

axes[1].set\_xlabel("Desire to Learn (Yes/No)")

axes[1].set\_ylabel("Count")

plt.tight\_layout()

plt.show()

# Step 5: Create Contingency Table and Perform Appropriate Test

contingency\_tables = {}

test\_results = {}

for age in df\_combined["age\_group"].unique():

table = pd.crosstab(df\_combined[df\_combined["age\_group"] == age]["rangeland\_experience"],

df\_combined[df\_combined["age\_group"] == age]["desire\_to\_learn"])

contingency\_tables[age] = table

# Check if more than 20% of expected frequencies are less than 5

chi2\_stat, p\_val, dof, expected = chi2\_contingency(table)

if (expected < 5).sum() > 0.2 \* expected.size:

# Use Fisher's Exact Test if the table is 2x2

if table.shape == (2, 2):

odds\_ratio, fisher\_p\_val = fisher\_exact(table)

test\_results[age] = {

"Test": "Fisher's Exact Test",

"Odds Ratio": odds\_ratio,

"P-value": fisher\_p\_val

}

else:

test\_results[age] = {

"Test": "Fisher's Exact Test",

"Error": "Fisher's Exact Test not applicable for tables larger than 2x2"

}

else:

# Use Chi-Square Test

test\_results[age] = {

"Test": "Chi-Square Test",

"Chi2 Statistic": chi2\_stat,

"P-value": p\_val,

"Degrees of Freedom": dof

}

# Display results

print("\n--- Test Results by Age Group ---")

for age, results in test\_results.items():

print(f"\nAge Group: {age}")

print("Contingency Table:\n", contingency\_tables[age])

for key, value in results.items():

print(f"{key}: {value}")

# Step 1: Filter necessary columns

df\_fixed = df\_backup.copy() # Use backup dataset

df\_fixed = df\_fixed[["age\_group", "what\_are\_your\_thoughts\_on\_prescribed\_burns?"]].dropna()

# Step 2: Rename the column for easier access

df\_fixed.rename(columns={"what\_are\_your\_thoughts\_on\_prescribed\_burns?": "prescribed\_burn\_opinion"}, inplace=True)

# Step 3: Create Contingency Table

contingency\_table = pd.crosstab(df\_fixed["age\_group"], df\_fixed["prescribed\_burn\_opinion"])

print("\nContingency Table:\n", contingency\_table)

# Step 4: Perform Chi-Square Test

chi2\_stat, p\_val, dof, expected = chi2\_contingency(contingency\_table)

# Display test results

print("\nChi-Square Test Results:")

print(f"Chi-Square Statistic: {chi2\_stat}")

print(f"P-value: {p\_val}")

print(f"Degrees of Freedom: {dof}")

# Interpretation

if p\_val < 0.05:

print(" Age group significantly influences opinions on prescribed burns.")

else:

print(" No significant relationship between age group and opinions on prescribed burns.")

# Step 5: Visualization

plt.figure(figsize=(10, 6))

sns.countplot(data=df\_fixed, x="prescribed\_burn\_opinion", hue="age\_group")

plt.title("Opinions on Prescribed Burns by Age Group")

plt.xlabel("Opinion on Prescribed Burns")

plt.ylabel("Count")

plt.legend(title="Age Group")

plt.xticks(rotation=15)

plt.show()

plt.figure(figsize=(8,6))

sns.countplot(data=df\_combined, x="learning\_gains", hue="age\_group")

plt.title("Learning Gains by Age Group")

plt.xlabel("Learning Gains (Yes/No)")

plt.ylabel("Count")

plt.legend(title="Age Group")

plt.show()