Descriptive Exercises

Garridos, Charlene P.

2024-02-02

Contents

1 Exercise 1: Analyzing Survey Data

1.1 Problem

A survey in which respondents were questioned about their preferred musical genres. The dataset `music_survey` has the following columns:

- `Respondent_ID`: Unique identifier for each respondent.
- `Age`: Respondent's age.
- `Gender`: Gender of the respondent (e.g., "Male," "Female," "Non-binary," "Prefer not to say").
- `Favorite_Music_Genre`: Favorite music genre of the respondent (e.g., "Rock," "Pop," "Hip-hop," "Jazz").

1.2 Dataset

```
# Read the CSV file
music_survey <- read.csv("music_survey.csv")

# View the data
head(music_survey)
```

```
##   Respondent_ID Age            Gender Favorite_Music_Genre
## 1             1  23              Male                 Rock
## 2             2  29 Prefer not to say                  Pop
## 3             3  26        Non-binary              Hip-hop
## 4             4  31        Non-binary              Hip-hop
## 5             5  32              Male                 Rock
## 6             6  35 Prefer not to say              Hip-hop
```

1.3 Frequency and Relative Frequency

```
Favorite_Music_Genre <- music_survey$Favorite_Music_Genre

# table() function to calculate frequencies
freq <- table(Favorite_Music_Genre)

# prop.table() function to calculate relative frequencies
relative_freq <- prop.table(freq)

# A data frame for the tables
result <- data.frame(Favorite_Music_Genre = names(freq), Frequency = as.vector(freq),
                     Relative_Frequency = as.vector(relative_freq))

result
```

```
##   Favorite_Music_Genre Frequency Relative_Frequency
## 1              Hip-hop        82             0.2050
## 2                 Jazz       100             0.2500
## 3                  Pop       113             0.2825
## 4                 Rock       105             0.2625
```

1.4 Questions with Answers

1. What is the most common favorite music genre among the survey respondents in terms of frequency?

   The most common favorite music genre is pop music. 113 respondents out of 400 love listening to pop music.

2. What is the most common favorite music genre among the survey respondents in terms of relative frequency?

   Pop music is the most popular genre among listeners. 28% of people enjoy listening to pop music.

3. Are there any noticeable gender or age-related patterns in the data?

   No, there are no noticeable gender- or age-related patterns in the data since, based on the instructions, we only analyze the distribution of favorite music genres among the survey respondents.

2 Exercise 2: Finding the Mode of Test Scores

2.1 Problem

Suppose you are a teacher, and you have collected test scores from a class of 30 students. The test scores range from 0 to 100, and you want to find the mode (most common score) to understand which score appears most frequently among your students.

2.2 Dataset

```
# Create dataset
test_scores <- c(85, 92, 78, 95, 85, 90, 78, 92, 88, 75,
                 92, 85, 88, 90, 75, 78, 95, 92, 85, 75,
                 88, 92, 75, 85, 90, 78, 88, 75, 85, 92)
```

2.3 Mode and Frequency

```
# Use the table() function to calculate frequencies
scores_freq <- table(test_scores)
scores_freq
```

```
## test_scores
## 75 78 85 88 90 92 95
##  5  4  6  4  3  6  2
```

```
# Calculate the mode (most frequent test score) from the given dataset.
scores_mode <- names(scores_freq[scores_freq == max(scores_freq)])
scores_mode
```

```
## [1] "85" "92"
```

Determine how many times the mode score appears in the dataset

```
# Count how many times the mode appears
mode_count <- sum(scores_freq[scores_freq == max(scores_freq)])

# Displays  how many times the mode score appears in the dataset
cat("The number of times the mode appears for both \"85\" and \"92\" is", mode_count, "\n")
```

```
## The number of times the mode appears for both "85" and "92" is 12
```

Display the mode score and its frequency.

```
# Display the mode score and its frequency.
cat("The mode of the test scores is bimodal:", scores_mode, "and the frequency is:",
    scores_freq, "\n")
```

```
## The mode of the test scores is bimodal: 85 92 and the frequency is: 5 4 6 4 3 6 2
```

3 Exercise 3: Employee Preferences

3.1 Problem

Suppose you have surveyed 200 employees in a company to understand their preferences for different office locations based on their departments and job roles.

You collected the following data:

- Department (HR, IT, Sales, Marketing)
- Job Role (Manager, Analyst, Assistant)
- Preferred Office Location (Downtown, Suburb, Remote)

Here's a summary of the data:

- Among HR employees:

    30 are Managers (10 prefer Downtown, 10 prefer Suburb, 10 prefer Remote)

    60 are Analysts (20 prefer Downtown, 20 prefer Suburb, 20 prefer Remote)

    40 are Assistants (15 prefer Downtown, 10 prefer Suburb, 15 prefer Remote)

- Among IT employees:

    25 are Managers (8 prefer Downtown, 10 prefer Suburb, 7 prefer Remote)

    50 are Analysts (15 prefer Downtown, 20 prefer Suburb, 15 prefer Remote)

    35 are Assistants (10 prefer Downtown, 10 prefer Suburb, 15 prefer Remote)

- Among Sales employees:

20 are Managers (5 prefer Downtown, 10 prefer Suburb, 5 prefer Remote)

40 are Analysts (15 prefer Downtown, 10 prefer Suburb, 15 prefer Remote)

30 are Assistants (10 prefer Downtown, 5 prefer Suburb, 15 prefer Remote)

- Among Marketing employees:

  15 are Managers (5 prefer Downtown, 5 prefer Suburb, 5 prefer Remote)

  30 are Analysts (10 prefer Downtown, 10 prefer Suburb, 10 prefer Remote)

  20 are Assistants (8 prefer Downtown, 5 prefer Suburb, 7 prefer Remote)

3.2 Dataset

```
Employees <- data.frame(
  Department = rep(c("HR", "IT", "Sales", "Marketing"), each = 9),
  Job_Role = rep(c("Manager", "Analyst", "Assistant"), times = 4),
  Pref.Office_Location = rep(c("Downtown", "Suburb", "Remote"), each = 3),
  Count = c(10, 20, 15, 10, 20, 10, 10, 20, 15,
            8, 15, 10, 10, 20, 10, 7, 15, 15,
            5, 15, 10, 10, 10, 5, 5, 15, 15,
            5, 10, 8, 5, 10, 5, 5, 10, 7)
)
Employees
```

```
##    Department  Job_Role Pref.Office_Location Count
## 1          HR   Manager             Downtown    10
## 2          HR   Analyst             Downtown    20
## 3          HR Assistant             Downtown    15
## 4          HR   Manager               Suburb    10
## 5          HR   Analyst               Suburb    20
## 6          HR Assistant               Suburb    10
## 7          HR   Manager               Remote    10
## 8          HR   Analyst               Remote    20
## 9          HR Assistant               Remote    15
## 10         IT   Manager             Downtown     8
## 11         IT   Analyst             Downtown    15
## 12         IT Assistant             Downtown    10
## 13         IT   Manager               Suburb    10
## 14         IT   Analyst               Suburb    20
## 15         IT Assistant               Suburb    10
## 16         IT   Manager               Remote     7
## 17         IT   Analyst               Remote    15
## 18         IT Assistant               Remote    15
## 19      Sales   Manager             Downtown     5
## 20      Sales   Analyst             Downtown    15
## 21      Sales Assistant             Downtown    10
## 22      Sales   Manager               Suburb    10
## 23      Sales   Analyst               Suburb    10
## 24      Sales Assistant               Suburb     5
## 25      Sales   Manager               Remote     5
```

```
## 26      Sales   Analyst              Remote    15
## 27      Sales Assistant              Remote    15
## 28  Marketing   Manager            Downtown     5
## 29  Marketing   Analyst            Downtown    10
## 30  Marketing Assistant            Downtown     8
## 31  Marketing   Manager              Suburb     5
## 32  Marketing   Analyst              Suburb    10
## 33  Marketing Assistant              Suburb     5
## 34  Marketing   Manager              Remote     5
## 35  Marketing   Analyst              Remote    10
## 36  Marketing Assistant              Remote     7
```

3.3 Contingency Table

```
# Create a contingency table using xtabs
contingency_table <- xtabs(Count ~ Employees$Department + Employees$Job_Role
                           + Employees$Pref.Office_Location, data = Employees)
contingency_table
```

```
## , , Employees$Pref.Office_Location = Downtown
##
##                     Employees$Job_Role
## Employees$Department Analyst Assistant Manager
##            HR             20        15      10
##            IT             15        10       8
##            Marketing      10         8       5
##            Sales          15        10       5
##
## , , Employees$Pref.Office_Location = Remote
##
##                     Employees$Job_Role
## Employees$Department Analyst Assistant Manager
##            HR             20        15      10
##            IT             15        15       7
##            Marketing      10         7       5
##            Sales          15        15       5
##
## , , Employees$Pref.Office_Location = Suburb
##
##                     Employees$Job_Role
## Employees$Department Analyst Assistant Manager
##            HR             20        10      10
##            IT             20        10      10
##            Marketing      10         5       5
##            Sales          10         5      10
```

3.4 Proportional Table

Calculate the column-wise proportional table to understand the distribution of Preferred Office Location by Department and Job Role.

```
# Calculate the proportional table
proportional_table1 <- prop.table(xtabs(Count ~ Employees$Department + Employees$Job_Role
                          + Employees$Pref.Office_Location, data = Employees),
                          margin = 2) * 100  # Margin = 2 for column-wise percentages

# Display the proportional table
proportional_table1
```

```
## , , Employees$Pref.Office_Location = Downtown
##
##                       Employees$Job_Role
## Employees$Department   Analyst Assistant   Manager
##            HR        11.111111 12.000000 11.111111
##            IT         8.333333  8.000000  8.888889
##            Marketing  5.555556  6.400000  5.555556
##            Sales      8.333333  8.000000  5.555556
##
## , , Employees$Pref.Office_Location = Remote
##
##                       Employees$Job_Role
## Employees$Department   Analyst Assistant   Manager
##            HR        11.111111 12.000000 11.111111
##            IT         8.333333 12.000000  7.777778
##            Marketing  5.555556  5.600000  5.555556
##            Sales      8.333333 12.000000  5.555556
##
## , , Employees$Pref.Office_Location = Suburb
##
##                       Employees$Job_Role
## Employees$Department   Analyst Assistant   Manager
##            HR        11.111111  8.000000 11.111111
##            IT        11.111111  8.000000 11.111111
##            Marketing  5.555556  4.000000  5.555556
##            Sales      5.555556  4.000000 11.111111
```

Calculate the cell-wise proportional table to express the proportion of each Preferred Office Location within each cell (Department and Job Role).

```
# Calculate the proportional table
proportional_table2 <- prop.table(xtabs(Count ~ Employees$Department + Employees$Job_Role
                          + Employees$Pref.Office_Location, data = Employees),
                          margin = c(1,2)) * 100  # Margin = c(1,2) for cell-wise percentages

# Display the proportional table
proportional_table2
```

```
## , , Employees$Pref.Office_Location = Downtown
##
```

```
##                       Employees$Job_Role
## Employees$Department  Analyst Assistant  Manager
##            HR        33.33333  37.50000 33.33333
##            IT        30.00000  28.57143 32.00000
##            Marketing 33.33333  40.00000 33.33333
##            Sales     37.50000  33.33333 25.00000
##
## , , Employees$Pref.Office_Location = Remote
##
##                       Employees$Job_Role
## Employees$Department  Analyst Assistant  Manager
##            HR        33.33333  37.50000 33.33333
##            IT        30.00000  42.85714 28.00000
##            Marketing 33.33333  35.00000 33.33333
##            Sales     37.50000  50.00000 25.00000
##
## , , Employees$Pref.Office_Location = Suburb
##
##                       Employees$Job_Role
## Employees$Department  Analyst Assistant  Manager
##            HR        33.33333  25.00000 33.33333
##            IT        40.00000  28.57143 40.00000
##            Marketing 33.33333  25.00000 33.33333
##            Sales     25.00000  16.66667 50.00000
```

Calculate the row-wise proportional table to see the distribution of Preferred Office Location within each Department and Job Role, expressed as percentages.

```
# Calculate the proportional table
proportional_table3 <- prop.table(xtabs(Count ~ Employees$Department + Employees$Job_Role
                          + Employees$Pref.Office_Location, data = Employees),
                          margin = 1) * 100  # Margin = 1 for row-wise percentages

# Display the proportional table
proportional_table3
```

```
## , , Employees$Pref.Office_Location = Downtown
##
##                       Employees$Job_Role
## Employees$Department   Analyst Assistant   Manager
##            HR        15.384615 11.538462  7.692308
##            IT        13.636364  9.090909  7.272727
##            Marketing 15.384615 12.307692  7.692308
##            Sales     16.666667 11.111111  5.555556
##
## , , Employees$Pref.Office_Location = Remote
##
##                       Employees$Job_Role
## Employees$Department   Analyst Assistant   Manager
##            HR        15.384615 11.538462  7.692308
##            IT        13.636364 13.636364  6.363636
##            Marketing 15.384615 10.769231  7.692308
##            Sales     16.666667 16.666667  5.555556
```

```
##
## , , Employees$Pref.Office_Location = Suburb
##
##                      Employees$Job_Role
## Employees$Department   Analyst Assistant   Manager
##            HR        15.384615  7.692308  7.692308
##            IT        18.181818  9.090909  9.090909
##            Marketing 15.384615  7.692308  7.692308
##            Sales     11.111111  5.555556 11.111111
```

3.5 Questions with Answers

1. Which Department and Job Role combination has the highest percentage of employees preferring Downtown as their office location?

   The Department and Job Role combination that has the highest percentage of employees preferring Downtown as their office location is in the Sales department, where Analysts have the highest percentage with 16.67%.

2. What percentage of HR Managers prefer Remote as their office location?

   The percentage of HR Managers preferring Remote as their office location is 7.69%.