Correlation Analysis (20 Categorical Pairs)

Anahit Baghdasaryan

2023-12-04

Correlation Analysis

1 Contingency table

1.1 Definition:

Contingency tables (also called crosstabs or two-ways tables) are used in statistics to summarize the relationship between categorical variables. A contingency table is a special type of frequency distribution table, where two variables are shown simultaneously.

2 Pearson χ^2 test

2.1 Definition

Pearson's χ^2 test is a statistical test applied to sets of categorical data to evaluate how likely it is that any observed difference between the sets arose by chance. It is the most widely used of many χ^2 tests.

It tests a null hypothesis stating that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical

3 Coefficients using χ^2 statistic

3.1 Contingency coefficient C

3.1.1 Definition Originally, the Pearson's contingency coefficient is calculated as:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

with n being the total number of observations. However, there is another option to correct this contingency coefficient as:

$$C_{\text{corr}} = \frac{C}{C_{\max}} = \sqrt{\frac{\min(k,l)}{\min(k,l)-1}}\sqrt{\frac{\chi^2}{\chi^2+n}}$$

with

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

and

$$C_{\max} = \sqrt{\frac{\min(k,l)-1}{\min(k,l)}}$$

where k and l are the number of categories of the two variables being compared.

3.2 Cramer's V

3.2.1 Definition

For a kxl contingency table, n(min(k,l)-1) is the maximal value of the χ^2 statistic, dividing χ^2 by the maximal value leads to a scaled version with maximal value 1. This idea is used by Cramer's V as follow:

$$V = \frac{x^2}{n(\min(k,l)-1)}$$

Data Preprocessing

```
data <- read_csv("data_tidy.csv")
```

```
## Rows: 95 Columns: 59
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (20): timestamp, gender, age, academic_level, field_of_study, university...
## dbl (39): id, study_env_Campus Common Spaces, study_env_Classroom, study_env...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(data)
```

```
## # A tibble: 6 x 59
##      id study_~1 study~2 study~3 study~4 study~5 study~6 study~7 backg~8 backg~9
##   <dbl>    <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     1        0       0       0       1       1       0       0       1       0
## 2     2        1       0       0       0       1       0       0       1       0
## 3     3        0       0       0       1       1       0       0       1       0
## 4     4        0       1       1       0       1       0       0       1       1
## 5     5        0       0       0       1       1       1       0       1       1
## 6     6        0       0       0       1       0       0       0       0       0
## # ... with 49 more variables:
## #   `background_noise_Prefer a bustling environment` <dbl>,
## #   `background_noise_Prefer complete silence` <dbl>,
## #   study_time_Afternoon <dbl>, `study_time_Early morning` <dbl>,
## #   study_time_Evening <dbl>, `study_time_Late morning` <dbl>,
## #   study_time_Night <dbl>, `study_time_No specific preference` <dbl>,
## #   `resources_Interactive simulations or applications` <dbl>, ...
```

```
columns_to_exclude <- c("id", "timestamp", "field_of_study", "university", "gpa")

# List of columns to convert to categorical
columns_to_convert <- setdiff(names(data), columns_to_exclude)

# Convert selected columns to categorical
data[columns_to_convert] <- lapply(data[columns_to_convert], as.factor)
```

Attendance and Extracurricular Activities

```
# Attendance and Extracurricular Activities
# Create a contingency table
attendance_extracurricular <- table(data$attendance,
                           data$extracurricular_activities)

# Display the contingency table
print(attendance_extracurricular)
```

```
##
##                          No Yes
##   Occasionally            4   5
##   Rarely                  2   0
##   Yes, always            28  21
##   Yes, most of the time  17  18
```

```
# Perform a chi-squared test of independence
chi_squared_attendance_extracurricular <- chisq.test(attendance_extracurricular)
```

```
## Warning in chisq.test(attendance_extracurricular): Chi-squared approximation may
## be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_attendance_extracurricular)
```

```
##
##  Pearson's Chi-squared test
##
## data:  attendance_extracurricular
## X-squared = 2.6382, df = 3, p-value = 0.4508
```

```
# Contingency coefficient C
ContCoef(data$attendance, data$extracurricular_activities, correct = TRUE)
```

```
## [1] 0.2324665
```

```
#Cramer's V and more
assocstats(xtabs(~data$attendance + data$extracurricular_activities))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 3.3997  3  0.33401
## Pearson          2.6382  3  0.45083
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.164
## Cramer's V        : 0.167
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between attendance and extracurricular activities in the given dataset. The contingency coefficient and Cramer's V values reinforce the notion of a weak association. It's important to note that the warning about the chi-squared approximation being incorrect may indicate that the sample size is small, and caution should be exercised in interpreting the results. If possible, in the future, we will consider a larger sample size for more reliable conclusions.

Academic Level and Eating Frequency

```
#Academic Level and Eating Frequency
# Create a contingency table
contingency_academic_eat <- table(data$academic_level,
                                  data$eat_freq)

# Display the contingency table
print(contingency_academic_eat)
```

```
##
##                   Four or more times I don't have regular meal times One time
##   Freshman                          1                                3        0
##   Graduate student                  0                                4        0
##   Junior                            0                                4        0
##   Senior                            1                                7        1
##   Sophomore                         2                                2        0
##
##                   Three times (breakfast, lunch, dinner) Two times
##   Freshman                                             4         4
##   Graduate student                                     7         3
##   Junior                                               5         8
##   Senior                                              11        13
##   Sophomore                                            3        12
```

```
# Perform a chi-squared test of independence
chi_squared_academic_eat <- chisq.test(contingency_academic_eat)
```

```
## Warning in chisq.test(contingency_academic_eat): Chi-squared approximation may
## be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_academic_eat)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_academic_eat
## X-squared = 13.963, df = 16, p-value = 0.6014
```

```
# Contingency coefficient C
ContCoef(data$academic_level, data$eat_freq, correct = TRUE)
```

```
## [1] 0.4002303
```

```
#Cramer's V and more
assocstats(xtabs(~data$academic_level + data$eat_freq))
```

```
##                      X^2 df P(> X^2)
## Likelihood Ratio 15.229 16  0.50796
## Pearson          13.963 16  0.60144
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.358
## Cramer's V        : 0.192
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between academic level and eating frequency in the given dataset. The contingency coefficient and Cramer's V values suggest a moderate association, but the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Attendance and Gender

```
#Attendance and Gender
# Create a contingency table
contingency_attendance_gender <- table(data$attendance,
                                       data$gender)

# Display the contingency table
print(contingency_attendance_gender)
```

```
##
##                         Female Male
##   Occasionally               5    4
##   Rarely                     1    1
##   Yes, always               37   12
##   Yes, most of the time     22   13
```

```
# Perform a chi-squared test of independence
chi_squared_attendance_gender <- chisq.test(contingency_attendance_gender)
```

```
## Warning in chisq.test(contingency_attendance_gender): Chi-squared approximation
## may be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_attendance_gender)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_attendance_gender
## X-squared = 2.6447, df = 3, p-value = 0.4497
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$attendance,
                                            data$gender, correct = TRUE))
```

```
## Contingency coefficient C: 0.2327461
```

```
#Cramer's V and more
assocstats(xtabs(~data$attendance + data$gender))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 2.6240  3   0.4533
## Pearson          2.6447  3   0.4497
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.165
## Cramer's V        : 0.167
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between attendance and gender in the given dataset. The contingency coefficient and Cramer's V values suggest a relatively weak association, and the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Eating Habits and Caffeine Consumption

```
#Eating Habits and Caffeine Consumption
# Create a contingency table
contingency_eating_caffeine <- table(data$eating_habits,
                                     data$caffeine_freq)

# Display the contingency table
print(contingency_eating_caffeine)
```

```
##
##                       Daily I don't consume caffeinated beverages
##   Healthy                 1                                     2
##   Neutral                13                                     0
##   Somewhat healthy        7                                     1
##   Somewhat unhealthy      9                                     1
##   Unhealthy               0                                     0
##
##                       Multiple times a day Occasionally Rarely or never
##   Healthy                                1            1               0
##   Neutral                               13            5               2
##   Somewhat healthy                      10            5               4
##   Somewhat unhealthy                     7            1               3
##   Unhealthy                              6            2               1
```

```
# Perform a chi-squared test of independence
chi_squared_eating_caffeine <- chisq.test(contingency_eating_caffeine)
```

```
## Warning in chisq.test(contingency_eating_caffeine): Chi-squared approximation
## may be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_eating_caffeine)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_eating_caffeine
## X-squared = 28.284, df = 16, p-value = 0.02924
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:",ContCoef(data$eating_habits,
                                           data$caffeine_freq, correct = TRUE))
```

```
## Contingency coefficient C: 0.5355131
```

```
#Cramer's V and more
assocstats(xtabs(~ data$eating_habits + data$caffeine_freq))
```

```
##                      X^2 df P(> X^2)
## Likelihood Ratio 23.816 16 0.093592
## Pearson          28.284 16 0.029239
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.479
## Cramer's V        : 0.273
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is a significant association between eating habits and caffeine consumption in the given dataset. The contingency coefficient and Cramer's V values suggest a moderate to strong association, and the p-value is less than 0.05, indicating that we have enough evidence to reject the null hypothesis of independence.

Library Study Environment and Age

```
#Library Study Environment and Age
# Create a contingency table
contingency_library_age <- table(data$study_env_Library, data$age)

# Display the contingency table
print(contingency_library_age)
```

```
##
##     18-24 25-30 31-40 Under 18
##   0    43     1     1        0
##   1    49     0     0        1
```

```
# Perform a chi-squared test of independence
chi_squared_library_age <- chisq.test(contingency_library_age)
```

```
## Warning in chisq.test(contingency_library_age): Chi-squared approximation may be
## incorrect
```

```
# Display the chi-squared test result
print(chi_squared_library_age)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_library_age
## X-squared = 3.1368, df = 3, p-value = 0.371
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:",ContCoef(data$study_env_Library,
                                           data$age, correct = TRUE))
```

```
## Contingency coefficient C: 0.2528395
```

```
#Cramer's V and more
assocstats(xtabs(~data$study_env_Library + data$age))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 4.2872  3  0.23208
## Pearson          3.1368  3  0.37101
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.179
## Cramer's V        : 0.182
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between the library study environment and age in the given dataset. The contingency coefficient and Cramer's V values suggest a relatively weak association, and the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Coffee Consumption and Caffeine Frequency

```
#Coffee Consumption and Caffeine Frequency
# Create a contingency table
contingency_coffee_caffeine <- table(data$consumed_beverages_Coffee,
                                     data$caffeine_freq)

# Display the contingency table
print(contingency_coffee_caffeine)
```

```
##
##     Daily I don't consume caffeinated beverages Multiple times a day
##   0     6                                     4                    5
##   1    24                                     0                   32
##
##     Occasionally Rarely or never
##   0           10               9
##   1            4               1
```

```
# Perform a chi-squared test of independence
chi_squared_coffee_caffeine <- chisq.test(contingency_coffee_caffeine)
```

```
## Warning in chisq.test(contingency_coffee_caffeine): Chi-squared approximation
## may be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_coffee_caffeine)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_coffee_caffeine
## X-squared = 38.946, df = 4, p-value = 7.146e-08
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:",ContCoef(data$consumed_beverages_Coffee,
                                           data$caffeine_freq, correct = TRUE))
```

```
## Contingency coefficient C: 0.7625758
```

```
#Cramer's V and more
assocstats(xtabs(~ data$consumed_beverages_Coffee + data$caffeine_freq))
```

```
##                     X^2 df   P(> X^2)
## Likelihood Ratio 41.334  4 2.2925e-08
## Pearson          38.946  4 7.1464e-08
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.539
## Cramer's V        : 0.64
```

Overall Interpretation:

The chi-squared test and associated measures suggest a highly significant association between coffee consumption and caffeine frequency in the given dataset. The contingency coefficient and Cramer's V values suggest a strong association, and the p-value is extremely small, indicating strong evidence to reject the null hypothesis of independence.

Tea Consumption and Caffeine Frequency

```
#Tea Consumption and Caffeine Frequency
# Create a contingency table
contingency_tea_caffeine <- table(data$consumed_beverages_Tea,
                                  data$caffeine_freq)

# Display the contingency table
print(contingency_tea_caffeine)
```

```
##
##     Daily I don't consume caffeinated beverages Multiple times a day
##   0    12                                     2                   19
##   1    18                                     2                   18
##
##     Occasionally Rarely or never
##   0            6               7
##   1            8               3
```

```
# Perform a chi-squared test of independence
chi_squared_tea_caffeine <- chisq.test(contingency_tea_caffeine)
```

```
## Warning in chisq.test(contingency_tea_caffeine): Chi-squared approximation may
## be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_tea_caffeine)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_tea_caffeine
## X-squared = 3.021, df = 4, p-value = 0.5543
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:",ContCoef(data$consumed_beverages_Tea,
                                           data$caffeine_freq, correct = TRUE))
```

```
## Contingency coefficient C: 0.2482744
```

```
#Cramer's V and more
assocstats(xtabs(~ data$consumed_beverages_Tea + data$caffeine_freq))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 3.0728  4  0.54572
## Pearson          3.0210  4  0.55431
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.176
## Cramer's V        : 0.178
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between tea consumption and caffeine frequency in the given dataset. The contingency coefficient and Cramer's V values suggest a relatively weak association, and the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Night Study Time and Caffeine Frequency

```r
#Night Study Time and Caffeine Frequency
# Create a contingency table
contingency_night_caffeine <- table(data$study_time_Night, data$caffeine_freq)

# Display the contingency table
print(contingency_night_caffeine)
```

```
##
##     Daily I don't consume caffeinated beverages Multiple times a day
##   0    15                                       1                   20
##   1    15                                       3                   17
##
##     Occasionally Rarely or never
##   0            9               4
##   1            5               6
```

```r
# Perform a chi-squared test of independence
chi_squared_night_caffeine <- chisq.test(contingency_night_caffeine)
```

```
## Warning in chisq.test(contingency_night_caffeine): Chi-squared approximation may
## be incorrect
```

```r
# Display the chi-squared test result
print(chi_squared_night_caffeine)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_night_caffeine
## X-squared = 2.6941, df = 4, p-value = 0.6103
```

```r
# Contingency coefficient C
cat( "Contingency coefficient C:",ContCoef(data$study_time_Night,
                                           data$caffeine_freq, correct = TRUE))
```

```
## Contingency coefficient C: 0.2348463
```

```r
#Cramer's V and more
assocstats(xtabs(~ data$study_time_Night + data$caffeine_freq))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 2.7569  4  0.59929
## Pearson          2.6941  4  0.61026
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.166
## Cramer's V        : 0.168
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between night study time and caffeine frequency in the given dataset. The contingency coefficient and Cramer's V values suggest a relatively weak association, and the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Exercise for Stress Management and Gender

```
#Exercise for Stress Management and Gender
# Create a contingency table
contingency_exercise_gender <- table(data$manage_stress_Exercise, data$gender)

# Display the contingency table
print(contingency_exercise_gender)
```

```
##
##     Female Male
##   0     57   23
##   1      8    7
```

```
# Perform a chi-squared test of independence
chi_squared_exercise_gender <- chisq.test(contingency_exercise_gender)
```

```
## Warning in chisq.test(contingency_exercise_gender): Chi-squared approximation
## may be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_exercise_gender)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  contingency_exercise_gender
## X-squared = 1.139, df = 1, p-value = 0.2859
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$manage_stress_Exercise,
                                            data$gender, correct = TRUE))
```

```
## Contingency coefficient C: 0.1968329
```

```
#Cramer's V and more
assocstats(xtabs(~data$manage_stress_Exercise + data$gender))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 1.7830  1  0.18178
## Pearson          1.8767  1  0.17071
##
## Phi-Coefficient   : 0.141
## Contingency Coeff.: 0.139
## Cramer's V        : 0.141
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between using exercise for stress management and gender in the given dataset. The contingency coefficient and Cramer's V values suggest a relatively weak association, and the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Night Study Time and Academic Level

```
#Night Study Time and Academic Level
# Create a contingency table
contingency_night_academic <- table(data$study_time_Night, data$academic_level)

# Display the contingency table
print(contingency_night_academic)
```

```
##
##     Freshman Graduate student Junior Senior Sophomore
##   0        6                7      8     17        11
##   1        6                7      9     16         8
```

```
# Perform a chi-squared test of independence
chi_squared_night_academic <- chisq.test(contingency_night_academic)

# Display the chi-squared test result
print(chi_squared_night_academic)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_night_academic
## X-squared = 0.46854, df = 4, p-value = 0.9765
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$study_time_Night,
                                            data$academic_level, correct = TRUE))
```

```
## Contingency coefficient C: 0.09907378
```

```
#Cramer's V and more
assocstats(xtabs(~data$study_time_Night + data$academic_level))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 0.47008  4  0.97635
## Pearson          0.46854  4  0.97649
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.07
## Cramer's V        : 0.07
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between night study time and academic level in the given dataset. The contingency coefficient and Cramer's V values suggest a very weak association, and the p-value is much greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Evening Study Time and Eating Habits

```
#Evening Study Time and Eating Habits
# Create a contingency table
contingency_evening_eating <- table(data$study_time_Evening, data$eating_habits)

# Display the contingency table
print(contingency_evening_eating)
```

```
##
##     Healthy Neutral Somewhat healthy Somewhat unhealthy Unhealthy
##   0       3       7                9                  7         5
##   1       2      26               18                 14         4
```

```
# Perform a chi-squared test of independence
chi_squared_evening_eating <- chisq.test(contingency_evening_eating)
```

```
## Warning in chisq.test(contingency_evening_eating): Chi-squared approximation may
## be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_evening_eating)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_evening_eating
## X-squared = 5.8234, df = 4, p-value = 0.2127
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$study_time_Evening,
                                            data$eating_habits, correct = TRUE))
```

```
## Contingency coefficient C: 0.3398764
```

```
#Cramer's V and more
assocstats(xtabs(~data$study_time_Evening + data$eating_habits))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 5.6858  4  0.22387
## Pearson          5.8234  4  0.21273
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.24
## Cramer's V        : 0.248
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between evening study time and eating habits in the given dataset. The contingency coefficient and Cramer's V values suggest a moderate association, but the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Attendance and Future Plans

```r
# Attendance and Future Plans
# Create a contingency table
contingency_attendance_plans <- table(data$attendance, data$future_plans)

# Display the contingency table
print(contingency_attendance_plans)
```

```
##
##                         No, I do not plan to pursue further education Undecided
##   Occasionally                                                      1         3
##   Rarely                                                            0         1
##   Yes, always                                                       2        13
##   Yes, most of the time                                             5         5
##
##                         Yes, but after gaining work experience
##   Occasionally                                                5
##   Rarely                                                      1
##   Yes, always                                                16
##   Yes, most of the time                                      13
##
##                         Yes, immediately after graduation
##   Occasionally                                          0
##   Rarely                                                0
##   Yes, always                                          18
##   Yes, most of the time                                12
```

```r
# Perform a chi-squared test of independence
chi_squared_attendance_plans <- chisq.test(contingency_attendance_plans)
```

```
## Warning in chisq.test(contingency_attendance_plans): Chi-squared approximation
## may be incorrect
```

```r
# Display the chi-squared test result
print(chi_squared_attendance_plans)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_attendance_plans
## X-squared = 10.377, df = 9, p-value = 0.3208
```

```r
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$attendance,
                                            data$future_plans, correct = TRUE))
```

```
## Contingency coefficient C: 0.362354
```

```r
#Cramer's V and more
assocstats(xtabs(~data$attendance + data$future_plans))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 13.857  9  0.12751
## Pearson          10.377  9  0.32083
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.314
## Cramer's V        : 0.191
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between attendance and future plans in the given dataset. The contingency coefficient and Cramer's V values suggest a moderate association, but the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Employment Status and Attendance

```
#Employment Status and Attendance
# Create a contingency table
contingency_employment_attendance <- table(data$employment_status,
                                           data$attendance)

# Display the contingency table
print(contingency_employment_attendance)
```

```
##
##                       Occasionally Rarely Yes, always Yes, most of the time
##   Employed full-time             4      0           9                     4
##   Employed part-time             2      1          10                    11
##   Unemployed                     3      1          30                    20
```

```
# Perform a chi-squared test of independence
chi_squared_employment_attendance <-
  chisq.test(contingency_employment_attendance)
```

```
## Warning in chisq.test(contingency_employment_attendance): Chi-squared
## approximation may be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_employment_attendance)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_employment_attendance
## X-squared = 7.2858, df = 6, p-value = 0.2952
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$employment_status,
                                            data$attendance, correct = TRUE))
```

```
## Contingency coefficient C: 0.3268706
```

```
#Cramer's V and more
assocstats(xtabs(~data$employment_status + data$attendance))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 6.7567  6  0.34394
## Pearson          7.2858  6  0.29523
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.267
## Cramer's V        : 0.196
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between employment status and attendance in the given dataset. The contingency coefficient and Cramer's V values suggest a moderate association, but the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Extracurricular Activities and Average Sleep

```
#Extracurricular Activities and Average Sleep
# Create a contingency table
contingency_extracurricular_sleep <- table(data$extracurricular_activities,
                                           data$avg_sleep)

# Display the contingency table
print(contingency_extracurricular_sleep)
```

```
##
##       5 to 6 hours 7 to 8 hours Less than 5 hours More than 8 hours
##   No            32           16                 1                 2
##   Yes           21           17                 4                 2
```

```
# Perform a chi-squared test of independence
chi_squared_extracurricular_sleep <-
  chisq.test(contingency_extracurricular_sleep)
```

```
## Warning in chisq.test(contingency_extracurricular_sleep): Chi-squared
## approximation may be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_extracurricular_sleep)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_extracurricular_sleep
## X-squared = 3.6172, df = 3, p-value = 0.3059
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$extracurricular_activities,
                                            data$avg_sleep, correct = TRUE))
```

```
## Contingency coefficient C: 0.2708465
```

```
#Cramer's V and more
assocstats(xtabs(~data$extracurricular_activities + data$avg_sleep))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 3.7412  3  0.29080
## Pearson          3.6172  3  0.30588
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.192
## Cramer's V        : 0.195
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between extracurricular activities and average sleep in the given dataset. The contingency coefficient and Cramer's V values suggest a weak to moderate association, but the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Study Hours and Average Sleep

```
#Study Hours and Average Sleep
# Create a contingency table
contingency_hours_sleep <- table(data$hours_spend_studying, data$avg_sleep)

# Display the contingency table
print(contingency_hours_sleep)
```

```
##
##                      5 to 6 hours 7 to 8 hours Less than 5 hours
##   11 to 20 hours               17           12                 1
##   21 to 30 hours               20            7                 1
##   31 to 40 hours                7            4                 2
##   41 to 50 hours                2            3                 0
##   Less than 10 hours            6            7                 1
##   more than 51 hours            1            0                 0
##
##                      More than 8 hours
##   11 to 20 hours                     1
##   21 to 30 hours                     2
##   31 to 40 hours                     1
##   41 to 50 hours                     0
##   Less than 10 hours                 0
##   more than 51 hours                 0
```

```
# Perform a chi-squared test of independence
chi_squared_hours_sleep <- chisq.test(contingency_hours_sleep)
```

```
## Warning in chisq.test(contingency_hours_sleep): Chi-squared approximation may be
## incorrect
```

```
# Display the chi-squared test result
print(chi_squared_hours_sleep)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_hours_sleep
## X-squared = 10.002, df = 15, p-value = 0.8196
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$hours_spend_studying,
                                            data$avg_sleep, correct = TRUE))
```

```
## Contingency coefficient C: 0.3563817
```

```
#Cramer's V and more
assocstats(xtabs(~ data$hours_spend_studying + data$avg_sleep))
```

```
##                      X^2 df P(> X^2)
## Likelihood Ratio 10.596 15  0.78066
## Pearson          10.002 15  0.81961
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.309
## Cramer's V        : 0.187
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between study hours and average sleep in the given dataset. The contingency coefficient and Cramer's V values suggest a weak to moderate association, but the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Home Study Environment and Break Preferences

```
#Home Study Environment and Break Preferences
# Create a contingency table
contingency_home_breaks <- table(data$study_env_Home, data$breaks)

# Display the contingency table
print(contingency_home_breaks)
```

```
##
##     10 to 20 minutes 21 to 40 minutes 41 to one hour Less than 10 minutes
##   0               11                5              2                     0
##   1               29               29              9                     4
##
##     More than an hour
##   0                 1
##   1                 5
```

```
# Perform a chi-squared test of independence
chi_squared_home_breaks <- chisq.test(contingency_home_breaks)
```

```
## Warning in chisq.test(contingency_home_breaks): Chi-squared approximation may be
## incorrect
```

```
# Display the chi-squared test result
print(chi_squared_home_breaks)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_home_breaks
## X-squared = 3.0662, df = 4, p-value = 0.5468
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$study_env_Home,
                                            data$breaks, correct = TRUE))
```

```
## Contingency coefficient C: 0.2500678
```

```
#Cramer's V and more
assocstats(xtabs(~ data$study_env_Home + data$breaks))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 3.7902  4  0.43515
## Pearson          3.0662  4  0.54680
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.177
## Cramer's V        : 0.18
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between home study environment and break preferences in the given dataset. The contingency coefficient and Cramer's V values suggest a weak association, but the p-value is greater than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Library Study Environment and Break Preferences

```
#Library Study Environment and Break Preferences
# Create a contingency table
contingency_library_breaks <- table(data$study_env_Library, data$breaks)

# Display the contingency table
print(contingency_library_breaks)
```

```
##
##     10 to 20 minutes 21 to 40 minutes 41 to one hour Less than 10 minutes
##   0               18               16              5                     2
##   1               22               18              6                     2
##
##     More than an hour
##   0                 4
##   1                 2
```

```
# Perform a chi-squared test of independence
chi_squared_library_breaks <- chisq.test(contingency_library_breaks)
```

```
## Warning in chisq.test(contingency_library_breaks): Chi-squared approximation may
## be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_library_breaks)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_library_breaks
## X-squared = 1.0149, df = 4, p-value = 0.9075
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:",ContCoef(data$study_env_Library,
                                           data$breaks, correct = TRUE))
```

```
## Contingency coefficient C: 0.1453959
```

```
#Cramer's V and more
assocstats(xtabs(~ data$study_env_Library + data$breaks))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 1.0257  4  0.90587
## Pearson          1.0149  4  0.90753
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.103
## Cramer's V        : 0.103
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between the library study environment and break preferences in the given dataset. The contingency coefficient and Cramer's V values indicate a very weak association, and the p-value is much higher than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Lecture Notes Usage and Help from Professors

```
#Lecture Notes Usage and Help from Professors
# Create a contingency table
contingency_lecture_help <- table(data$`resources_Lecture notes`,
                                  data$help_from_professors)

# Display the contingency table
print(contingency_lecture_help)
```

```
##
##     Always Frequently Never Occasionally Rarely
##   0      1          3     3            7      9
##   1      3          8     3           33     25
```

```
# Perform a chi-squared test of independence
chi_squared_lecture_help <- chisq.test(contingency_lecture_help)
```

```
## Warning in chisq.test(contingency_lecture_help): Chi-squared approximation may
## be incorrect
```

```
# Display the chi-squared test result
print(chi_squared_lecture_help)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_lecture_help
## X-squared = 3.3087, df = 4, p-value = 0.5076
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:",
     ContCoef(data$`resources_Lecture notes`, data$help_from_professors,
              correct = TRUE))
```

```
## Contingency coefficient C: 0.2594463
```

```
#Cramer's V and more
assocstats(xtabs(~ data$`resources_Lecture notes` + data$help_from_professors))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 3.0598  4  0.54788
## Pearson          3.3087  4  0.50756
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.183
## Cramer's V        : 0.187
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between lecture notes usage and help from professors in the given dataset. The contingency coefficient and Cramer's V values indicate a weak association, and the p-value is much higher than 0.05, indicating that we do not have enough evidence to reject the null hypothesis of independence.

Night Study Time and Future Plans

```
#Night Study Time and Future Plans
# Create a contingency table
contingency_night_plans <- table(data$study_time_Night, data$future_plans)

# Display the contingency table
print(contingency_night_plans)
```

```
##
##     No, I do not plan to pursue further education Undecided
##   0                                             4        13
##   1                                             4         9
##
##     Yes, but after gaining work experience Yes, immediately after graduation
##   0                                     16                                16
##   1                                     19                                14
```

```
# Perform a chi-squared test of independence
chi_squared_night_plans <- chisq.test(contingency_night_plans)
```

```
## Warning in chisq.test(contingency_night_plans): Chi-squared approximation may be
## incorrect
```

```
# Display the chi-squared test result
print(chi_squared_night_plans)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_night_plans
## X-squared = 1.024, df = 3, p-value = 0.7954
```

```
# Contingency coefficient C
cat( "Contingency coefficient C:", ContCoef(data$study_time_Night,
                                            data$future_plans, correct = TRUE))
```

```
## Contingency coefficient C: 0.1460434
```

```
#Cramer's V and more
assocstats(xtabs(~ data$study_time_Night + data$future_plans))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 1.0275  3  0.79460
## Pearson          1.0240  3  0.79544
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.103
## Cramer's V        : 0.104
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between night study time and future plans in the given dataset. The contingency coefficient and Cramer's V values indicate a weak association, and the p-value is much higher than 0.05, suggesting that we do not have enough evidence to reject the null hypothesis of independence.

Home Stydy Environment and Physical Activity Frequency

```r
# Home Stydy Environment and Physical Activity Frequency
# Create a contingency table
contingency_home_activity <- table(data$study_env_Home,
                                   data$physical_activity_freq)

# Display the contingency table
print(contingency_home_activity)
```

```
##
##     Daily Never Once a week Rarely Several times a week
##   0     3     1           3      3                    9
##   1     6    10           9     28                   23
```

```r
# Perform a chi-squared test of independence
chi_squared_home_activity <- chisq.test(contingency_home_activity)
```

```
## Warning in chisq.test(contingency_home_activity): Chi-squared approximation may
## be incorrect
```

```r
# Display the chi-squared test result
print(chi_squared_home_activity)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_home_activity
## X-squared = 5.3905, df = 4, p-value = 0.2495
```

```r
# Contingency coefficient C
cat( "Contingency coefficient C:",
     ContCoef(data$study_env_Home, data$physical_activity_freq, correct = TRUE))
```

```
## Contingency coefficient C: 0.3277054
```

```r
#Cramer's V and more
assocstats(xtabs(~ data$study_env_Home + data$physical_activity_freq))
```

```
##                     X^2 df P(> X^2)
## Likelihood Ratio 5.6848  4  0.22396
## Pearson          5.3905  4  0.24952
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.232
## Cramer's V        : 0.238
```

Overall Interpretation:

The chi-squared test and associated measures suggest that there is no significant association between the home study environment and physical activity frequency in the given dataset. The contingency coefficient and Cramer's V values indicate a moderate association, and the p-value is greater than 0.05, suggesting that we do not have enough evidence to reject the null hypothesis of independence.

Conclusion

In summary, the majority of the analyses did not find significant associations, and where associations were detected, they tended to be weak. The interpretation should consider the context of the study and the dataset's characteristics. It's essential to recognize the limitations of the chi-squared test and association measures and interpret the results accordingly.