Data Analysis - The Test of Association

Student: Kevin Ivanov

The Claims to Test

Test to see whether there is an association between Level of Stress and Home Environment,

A) Which procedure is the correct procedure for analyzing this data? How do you know?

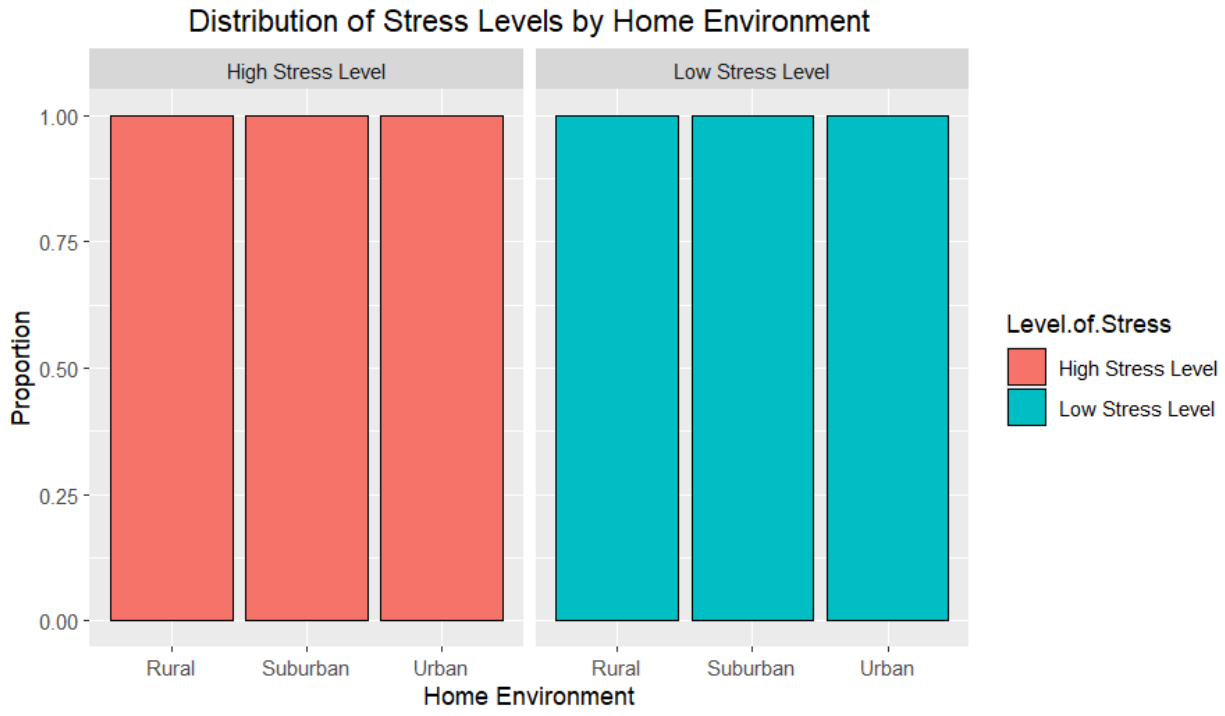
Given both variables are categorical, the appropriate procedure to test for an association between them is the Chi-Square Test of Independence. This test will help determine if there's a statistically significant association between the two categorical variables.

B) State the Null and the Alternative Hypothesis. Use appropriate notation.

Ho: Level of Stress is independent of Home Environment

H1: Level of Stress is not independent of Home Environment

C) [7.6.2] Construct a stacked bar chart with Level of Stress as the grouping variable.



D) Check whether the conditions (the assumptions we make in using the test) are satisfied. The population is all Pacific students.

Condition	Satisfied Yes/No	Justification
the observed cell counts are based on a random sample	Yes	The data was collected through a random sampling process, ensuring that the sample is representative of the Pacific students population, which reduces bias
every expected cell count is at least 5	Yes	All expected cell counts are above 5, which is necessary for the validity of the Chi-Square test

E) Compute the hypothesis test [13.10] using R. Copy the full output from R into the submission.

```
Total Observations in Table:  154

                          | LabData$Level.of.Stress
LabData$Home.Environment  | High Stress Level  |  Low Stress Level  |         Row Total |
--------------------------|--------------------|--------------------|-------------------|
                    Rural |                  7 |                  5 |                12 |
                          |              5.610 |              6.390 |                   |
                          |              0.344 |              0.302 |                   |
                          |             58.333% |             41.667% |             7.792% |
                          |              9.722% |              6.098% |                   |
                          |              4.545% |              3.247% |                   |
--------------------------|--------------------|--------------------|-------------------|
                 Suburban |                 47 |                 51 |                98 |
                          |             45.818 |             52.182 |                   |
                          |              0.030 |              0.027 |                   |
                          |             47.959% |             52.041% |            63.636% |
                          |             65.278% |             62.195% |                   |
                          |             30.519% |             33.117% |                   |
--------------------------|--------------------|--------------------|-------------------|
                    Urban |                 18 |                 26 |                44 |
                          |             20.571 |             23.429 |                   |
                          |              0.321 |              0.282 |                   |
                          |             40.909% |             59.091% |            28.571% |
                          |             25.000% |             31.707% |                   |
                          |             11.688% |             16.883% |                   |
--------------------------|--------------------|--------------------|-------------------|
             Column Total |                 72 |                 82 |               154 |
                          |             46.753% |             53.247% |                   |
--------------------------|--------------------|--------------------|-------------------|


Statistics for All Table Factors


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  1.307306     d.f. =  2     p =  0.5201423



       Minimum expected frequency: 5.61039
```

F) Give a statement about the value of the test statistic, the p-value, and the degrees of freedom.

Chi-Squared Statistic (χ2): 1.307306

Degrees of Freedom (d.f.): 2

P-value: 0.5201423

G) Give a statement either rejecting Ho or failing to reject Ho. Use a 5% significance level.

For 2 df, the critical value at a 95% confidence level = 5.991

Chi(1.307306) is less than the critical value from the table (5.991), we do not have sufficient evidence to reject the null hypothesis at the 5% significance level**.**

H) Give an interpretation of the results following class guidelines.

There is insufficient evidence to indicate that "Home Environment" and "Level of Stress" are related. So, we can affirm that random variation, rather than a real effect, might have been the basis for any observed differences in stress levels among subjects in various home environments

Code

```
ggplot(LabData, aes(x = Home.Environment, fill = Level.of.Stress)) + geom_bar(position = "fill", color = "black") + ggtitle("Distribution of Stress Levels by Home Environment") + xlab("Home Environment") + ylab("Proportion") + facet_grid(~Level.of.Stress) + theme(axis.line = element_blank(), plot.title = element_text(hjust = 0.5))
install.packages("gmodels")
result <- CrossTable(LabData$Home.Environment, data$Level.of.Stress, chisq = TRUE, format = "SPSS")
```