Homework 2

Please submit the solution in the form of R Markdown report, knitted into either of the available formats (HTML, pdf or Word). Provide all relevant code and output. Goal of this homework is to have you 1) familiarized with χ^2-test of independence for contingency tables; 2) measuring practical strength of dependence between categorical variables; 3) practice your R coding.

Problem #1

1. Code up your own *my.chisq.test*() function that will perform a χ^2-test. As a single argument, it should just take a contingency table of arbitrary size. As output, it should provide:

   - Calculated X^2 statistic.
   - p-value.

   Calculating the expected cell counts under H_0 hypothesis should constitute a critical part of your function definition. **Don't** use neither *chisq.test*(), nor *prop.test*(), nor any other "fancy cheat" R's built-in functions inside your function's definition.

2. For NYC Airbnb data set (*listings.csv* on Canvas), you would like to know whether there are differences in Airbnb room types offered in different NYC burrows. Proceed to formulate this in the form of a hypothesis test, as in:

   a. What variables are we interested in?
   b. What are the hypotheses?
   c. Print the contingency table. Under H_0 hypothesis, proceed to calculate expected counts for **two arbitrary** cells of the contingency table (simply for practice).
   d. Proceed to apply your *my.chisq.test*() and interpret the results. As a sanity check, also run R's built-in *chisq.test*() function on that same data, make sure the outputted X^2 and p-values match with those provided by *my.chisq.test*().
   e. In case you end up claiming that variables are not independent, proceed to make a few comments on **strength** of the relationship (as was done for Income & Happiness example in class).

Problem #2

From **Agresti** book, do exercises:

1. **11.84**
2. **11.9 (+ calculate all the expected cell counts, for practice)**
3. **11.16**

Problem #3.

1. For all three examples on the "X^2 Does **NOT** Measure **Strength** of Association" slide, proceed to
    a. Use the *my.chisq.test*() function you've defined previously in order to confirm the X^2 and p-values. **Hint**: Make sure to convert the %'es into counts first.
    b. Calculate the difference in proportion between males & females that attend religious services weekly. Calculate the risk ratio between males & females that attend religious services weekly.
    c. Based on your answers to parts $(a) - (b)$, as n increases, what do you notice with respect to statistical significance? Practical significance?
2. Do exercise **11.32** from **Agresti** book.