Exam 02: Takehome Portion

Name: _____

(20 points) This take-home exam is due at the start of class on Monday, 25 March. You must work on your own, but may use any static external references that you need. Note that I have uploaded an R notebook that you may find useful to compute the summary statistics below, but it is not obligatory to use R (but it is likely much easier) nor do you need to hand in any of the code. This sheet is all that I need.

General Tasks For this exam, you will need to collect two datasets: (1) a 2x2 contingency table and (2) data for a simple linear regression. For the contingency table, you should aim to have at least 25 counts and at least 3 counts in each cell. You are free to create whatever data you want; you are welcome to use something where the data are already counted for you, or you can create the counts yourself. You will make the regression data on the second page from your own experimentation.

1. Write down the contingency table that you generated, labeling the rows and columns with enough information along with a sentence or two (if needed) to explain the meaning of the data.

2. Consider the null-hypothesis that the rows and columns are independent. What is the G-score test statistic? What is the p-value? Would you reject the null at the 0.05 level?

3. Consider the null-hypothesis that the rows and columns are independent. What is the chi-squared test statistic? What is the p-value? Would you reject the null at the 0.05 level?

4. Below are ten world cities, all with a population over 1 million people. Write down your guess of their population size in millions; this is your **y**. Then, lookup the actual size in millions (round to the nearest million); this is your **x**.

- Tehran
- Lima
- Houston
- London
- Cairo
- Bogotá
- Seoul
- Osaka
- Moscow
- Paris

5. Run a linear regression with a slope and intercept. Write down the point estimates for both.

6. Is the test-statistic for the slope parameter significant at the 0.05 level?