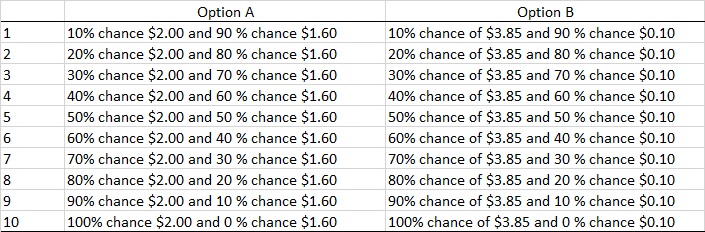
**Assignment 2**

**Problem 1: (50 pts): Analyzing data from a risk preference experiment.**

Risk preference (<https://www.sapling.com/8665868/difference-risk-averse-risk-neutral>) is one the most fundamental preferences in behavioral economics and finance.



In experimental economics, we usually use the Holt and Laury questionnaire (Holt and Laury, 2002) to measure individual risk preferences. mpl.txt file contains the data collected from a classroom experiment from EC406-2019 (Experimental Economics) using this questionnaire, each column separated by \t. For variable treatment, 1 is for control group, and 2 is for treatment group. Variables choice1-choice10 indicate whether one chooses A or B for each question. We can measure people’s risk preferences by observing the choice at which one switches from choosing A over to B. Switching between 1-4 means a person is risk-seeking, while switching between 5-10 means a person is risk-averse. Now please do the following:

1. Write to a file switch.txt, containing three variables: id, treatment, switch. id and treatment are the same as the original data, and switch is each individual’s A-B switching point.
2. Print to screen the average switching point for both control and treatment groups.
3. Print to screen the number of people who are risk-averse for both control and treatment groups.

**Problem 2: (50 pts): Sentiment analysis.**

Modern programming languages greatly facilitate a kind of data analysis method called text analysis. Sentiment analysis, where computer programs extract and analyze human sentiment from text data, is a type of text analysis. Researchers have been using sentiment analysis on social media content to predict the movement of stock prices (<http://www.sciencedirect.com/science/article/pii/S187775031100007X>).

lexicon.txt is a “dictionary” for words and their sentimental values (from <http://mpqa.cs.pitt.edu/>). Each line looks like this:

type=weaksubj len=1 word1=active pos1=adj stemmed1=n priorpolarity=positive

What is useful to you is word1= and priorpolarity=positive/negative. For example, the above line means the word “active” is positive.

review.txt contains randomly chosen 200 product reviews from Amazon. Please apply sentiment analysis on each review:

The base score is 0; whenever a “positive” word is encountered, the score is increased by 1; whenever a “negative” word is encountered, the score is decreased by 1.

As a result, print to screen the average sentiment score, and the number of positive/neutral/negative reviews (score >/=/< 0 respectively).