Dear Candidate:

Thank you for your application. This file consists of two task assignments. These assignments consist of research skills assessments that will help us determine fit in our economics research specialist position. These assessments are intended to test your ability to work independently, applied economic thinking, program, and find creative solutions to problems. If there are multiple approaches to a problem, state and justify your assumptions in your response and move on. **Please read the instructions for each assignment carefully before beginning your work.**

We will invite you for a Zoom interview shortly after your submission if we determine that you would be a good match for the position.

Please do not hesitate to contact us if you have any questions. We request that you complete this assignment by January 15, 11:59 pm ET.

Regards,

Prof. Apoorv Gupta

Prof. Jonathan Zinman

Task 1: Red Sox Ticket Prices

Time: Spend no more than 8 hours on this task

Files: Red_Sox.zip

The .zip archive contains a comma-separated data file organized by year, containing a list of all StubHub ticket orders for the Boston Red Sox. The file includes information about the characteristics of the game (e.g., opponent, time-of-day, day-of-week, etc.) and other relevant order characteristics (e.g., number of tickets in order, etc.)

We'd like you to use these data to produce your best answer to the following questions:

** How do the prices consumers pay for tickets change as the game date approaches (i.e., as the number of days between transaction date and game date declines)? How does this dynamic pattern change across years?

We'd like you not only to decide what you think is the answer but also to prepare a document arguing for that based on the evidence.

You should feel free to use whatever techniques you want. The goal here is not to show off hi-tech econometrics, but rather to show us how you think about data. Sometimes something as simple as a graph can do more for an argument than all the estimators in the world.

Also, you need not take the structure of data as written in stone. If there are ways you'd like to transform the variables or restructure the data, please feel free. Think of this as a small research project: you have a dataset, and you need to find the best answer you can to a real-world question. You also need to justify it to an outsider (in this case, us!).

You are welcome to use whatever tool you like for this, including Python, R, Stata, Matlab, or another program.

Task 2: Lottery Study

Time: No limit; complete as quickly as you can

Files: lottery_study.zip

Pages 5-8 list the definitions of variables constructed from responses to a recent survey about participation in lotteries. The **bold, red text** indicates the variable name. The accompanying dataset, lottery_study.dta, contains the resulting survey data.

Using this dataset, please write a script with commented, easy-to-read code that carries out the tasks below. Submit both the script and a PDF with your results. If you know Latex (or Lyx, Overleaf, etc.), please use that to write up your answers.
For tasks I.2-I.6, you can simply screenshot your output in Stata. Please export 1-2 presentation-quality figures and/or regression tables for each of the tasks II.1-II.2; you may screenshot any other supporting output. As you complete the tasks, feel free to transform the variables in the dataset in any manner you see fit.

PART I – Descriptive Data Analysis

1. Do any of the observations seem suspicious to you? If so, please provide a one-sentence justification and drop these observations from the analysis that follows. (Hint: Look for observations with implausible values.)
2. Report a tabulation of income and gender.
3. Report detailed summary statistics for age which include the mean, standard deviation, median, and 25th/75th percentiles. In Stata, display the sentence: "The mean age in the sample is *x*," where *x* is the mean you just reported.
4. Report summary statistics for monthly expenditures on lottery tickets as well as those expenditures as a share of monthly income. Generate a histogram of monthly lottery ticket expenditures.
5. Using the education variable, generate indicator variables that have a value of 1 if the respondent's years of education are greater than or equal to *y*, for *y* in {5, 10, 15, 20}, and have a value of 0 otherwise. In naming these variables, you can simply add the threshold value *y* to the end of the existing variable's name (e.g., education_20). You do not need to display anything in your results PDF for this task; the code is sufficient.
6. On average, how did changes in income from 2018 to 2019 vary with changes in lottery ticket expenditures? Does this differ from how much respondents imagine a change in their income would affect their lottery expenditures?

PART II – Unpacking the determinants and correlates of lottery expenditures

The questions in this part are less specific because their purpose is to give you a chance to illustrate your ability and willingness to take initiative and dive into a new dataset to explore high-level research questions.

1. Please explore the relationship between the monthly lottery expenditure variable and income. Write a paragraph to describe your findings, and include statistical tests (e.g., confidence intervals) for any relationships that you do or do not find. Do your findings confirm or contradict what you know or have heard about the lottery?
2. Please explore the relationship between the monthly lottery expenditure variable and the preference and bias proxy variables. Write two succinct paragraphs to describe your findings, and include statistical tests (e.g., confidence intervals) for any relationships that you do or do not find. In particular, does the data suggest that behavioral biases play a large or small role in people's lottery expenditures?

PART III – Short Answer

Please answer the following questions with no more than one paragraph.

A. After each Mega Millions lottery drawing, approximately 50% of sales from the previous drawing are added to what is called the "prize pool." From this total prize pool, a different share is added to a specific sub-pool for each prize. For example, 7% of the total prize pool might be allocated to the sub-pool for the second-largest prize, which is won when the player picks five numbers correctly but picks the "Mega Ball" incorrectly. Please find a reference that documents the share of the total prize pool that is added to each prize level's sub-pool by Mega Millions. Since Mega Millions occasionally changes its rules, any source you find from the past decade suffices. You might find documentation from the lottery itself or something from a governmental source since lotteries are run by state governments. (Note: California and other states have different Mega Millions regulations; any state you can find is fine for the purposes of this task.) Please do not spend more than 25 minutes on this task. If you are unable to find a source, please describe how you tried to find this information (e.g., specific Google search terms you used).
B. What is an interesting economics paper that you have recently read (published in a peer-reviewed journal in the past decade) and what did you find interesting about it? (You do not need to provide a full citation of the paper; the title and author(s) are sufficient.)

DEMOGRAPHIC VARIABLES

income
Annual household income in thousands of dollars. Responses are discrete, but you can treat them as continuous.

age
Age in years.

black
1 if Black, 0 otherwise.

hispanic
1 if Hispanic, 0 otherwise.

white
1 if White, 0 otherwise.

gender
1 if male, 2 if female, 3 otherwise.

marital
1 if married, 0 otherwise.

urban
1 if residing in a metropolitan area, 0 otherwise.

employment
1 if working as a paid employee or self-employed, 0 otherwise.

religion
1 if attending religious services at least once a month, 0 otherwise.

education
Years of education.

ideology
1 if extremely liberal, 2 if liberal, 3 if slightly liberal, 4 if moderate, 5 if slightly conservative, 6 if conservative, and 7 if extremely conservative.

state
State of residence.

SPENDING & INCOME EFFECTS VARIABLES

expend_total
Monthly lottery expenditures in dollars.

income_delta
Response to the question: “How much income did you earn in 2019 compared to 2018?” This variable reports the percentage change (e.g., a 2.5% increase). Responses are discrete, but you can treat them as continuous.

expend_delta
Response to the question: “How much money did you spend in total on lottery tickets in 2019 compared to 2018?” This variable reports the percentage change (e.g., a 2.5% increase). Responses are discrete, but you can treat them as continuous.

income_effects_delta_pct
Response to the question: “Imagine you got a raise and your income doubled. How do you think your lottery spending would change?” This variable reports the percentage change (e.g., a 2.5% increase). Responses are discrete, but you can treat them as continuous.

PREFERENCE & BIAS PROXY VARIABLES

risk_seeking
Response to the question: “In general, how willing or unwilling are you to take risks?” on a scale of -7 to -1, where -1 is “very unwilling” and -7 is “very willing.”

risk_aversion
Response to the question: “Which of the following statements comes closest to the amount of financial risk that you are willing to take when you save or make investments?”

- 1 if “Substantial financial risks expecting to earn substantial returns,”
- 2 if “Above-average financial risks expecting to earn above-average returns,”
- 3 if “Average financial risks expecting to earn average returns,” and
- 4 if “No financial risks”

seems_fun
Response to the question: “To what extent do you agree or disagree with the following statement: *For me, playing the lottery seems fun*” on a scale of -3 to 3, where -3 is “strongly disagree” and 3 is “strongly agree.”

enjoy_thinking
Response to the question: “To what extent do you agree or disagree with the following statement: I enjoy thinking about how life would be if I won the lottery” on a scale of -3 to 3, where -3 is “strongly disagree” and 3 is “strongly agree.”

self_control
Response to the question: “It can be hard to exercise self-control, and some people feel that there are things they do too much or too little—for example, exercise, save money, or eat junk food. Do you feel like you play the lottery too little, too much, or the right amount?” on a scale of -3 to 3, where -3 is “far too little,” 0 is “the right amount,” and 3 is “far too much.”

financial_literacy
The fraction of a series of five financial literacy questions that the respondent correctly answered.

financial_numeracy
The fraction of a series of three financial numeracy questions that the respondent correctly answered.

gamblers_fallacy
The fraction of the following questions that the respondent *incorrectly* answered: “For the next few questions, imagine flipping a coin that has a 50% chance of landing heads and a 50% chance of landing tails. Imagine that after eight flips, you observe the patterns described in the table below. What is the probability, in percent from 0-100, that the next flip is tails?

- tails-tails-tails-heads-tails-heads-heads-heads
- heads-heads-heads-heads-heads-heads-heads-heads
- heads-tails-heads-tails-tails-tails-tails-tails”

non_belief_lln
The mean absolute difference between the respondent’s answer and the correct answer to the following questions: “Now imagine starting over and flipping a coin 1000 times. What are the chances, in percent from 0-100, that the total number of heads will lie within the following ranges?

- Between 481 and 519 heads
- Between 450 and 550 heads
- Between 400 and 600 heads”

ev_miscalculation
The fraction of the following questions that the respondent *incorrectly* answered: “Now we are going to ask you how much people might win from different lotteries. For each lottery described below, please give us your best estimate of what percent (from 0-100) of the lottery revenues are returned to the winners.

- Tickets cost $1, and 1 out of every 10 tickets wins $10.
- Tickets cost $1, and 1 out of every 1,000 tickets wins $500.
- Tickets cost $1, 1 out of every 400,000,000 tickets wins $200,000,000, and 1 out of every 1,000 tickets wins $100.
- Tickets cost $1, and 1 out of every 300,000,000 tickets wins $200,000,000.”

overconfidence
For every dollar the respondent spends on lottery tickets in the long run, the amount in dollars more they expect that they would win compared to the average lottery player.

lottery_payout
Response to the question: “Think about the total amount of money spent on lottery tickets nationwide. What percent do you think is given out in prizes?” Responses are reported as a fraction (from 0 to 1).

happiness
On a scale of -10 to 10, the extent to which the respondent believes an additional $100,000 in lottery winnings increased the average life satisfaction of Swedish lottery winners in a recent study.