14.310x: Data Analysis for Social Scientists Moments of a Random Variable, Applications to Auctions, and Intro to Regression

Welcome to your fifth homework assignment! You will have about one week to work through the assignment. We encourage you to get an early start, particularly if you still feel you need more experience using R. We have provided this PDF copy of the assignment so that you can print and work through the assignment offline. You can also go online directly to complete the assignment. If you choose to work on the assignment using this PDF, please go back to the online platform to submit your answers based on the output produced.
Good luck :)!

In the first part of the problem set, we will delve more deeply into **auction theory**, which Sara introduced in lecture. We will demonstrate some auction theory properties by performing simulations of data. In these simulations, we will compare different schemes for auctions by varying the number of bidders and valuations. At an auction, **bidders** make offers to buy the goods, and a bidder's **valuation** is how much the bidder offers to pay for the good.

To start, try to understand the following R code. Run the code to test your understanding. We will start with the assumption that there are 2 bidders. We will simulate the auction 1000 times, resulting in 1000 valuations for these 2 bidders. Imagine that you are the person trying to sell a particular good, and that you are using R to figure out the perfect pricing and allocation scheme.

```
# Preliminaries
rm(list = ls())
setwd("/Users/hol/Dropbox (MIT)/2016 Fall/14.310x/")

# Uniform Valuations
number_of_bidders <- 2
number_of_simulations <- 1000

set.seed(1)
valuations1 <- matrix(runif(
  number_of_bidders*number_of_simulations, min=0, max=1),
  nrow = number_of_simulations)
```

1. Try to figure out what the set.seed() command is doing, and then answer the following true or false question:

   True or False: Even though we are simulating random numbers, the use of this set.seed() allows you to have the same valuations whenever you decide to re run this code *a*) True

b) False

2. Let us consider the **posted price example**. In lecture, we saw that the expected revenue when there are posted prices is given by $pPr(v_i \geq p)$ for at least one i, where p corresponds to the posted price and v_i for the valuation of the good of individual i. Then, the expected revenue is equal to: $p(1 - Pr(v_{(N)} < p)) = p(1 - F(p)^N)$.

   What is the optimal price in the case of two bidders, and a U[0,1] distribution for valuations?

   a) $\frac{1}{\sqrt[3]{2}}$

   b) $\frac{1}{\sqrt[3]{3}}$

   c) $\frac{1}{\sqrt{3}}$

   d) $\frac{1}{\sqrt{2}}$

3. Using the same scenario from Question 2, what is the expected revenue for the seller?

   a) The expected revenue of the posted price would be given by $\frac{2}{3\sqrt{3}}$

   b) The expected revenue of the posted price would be given by $\frac{2}{3\sqrt[3]{2}}$

   c) The expected revenue of the posted price would be given by $\frac{2}{3\sqrt[3]{3}}$

   d) The expected revenue of the posted price would be given by $\frac{3}{2\sqrt[3]{2}}$

4. Now, we will use the R code above to test whether these predictions hold.

   First, let's find the maximum valuation among the two bidders. Name the function in R that allows you to get this value. The function you use should return the maximum valuation when you run *funcName(valuations1)*.

   Please enter only the function name (what you typed for funcName) – no arguments or parentheses!

5. Take a look at the following R code that calculates the analytic solution to the expected revenue, and compare it with the one coming from the simulation.

```r
#Uniform Valuations
number_of_bidders <- 2
N <- number_of_bidders
V <- 10000

set.seed(5)
valuations <- matrix(runif(
  N*V, min = 0, max = 1),
  nrow = V)

maximum_valuation <- apply(valuations, 1, max)
optimal_price <- 1/((N+1)^(1/N))
expected_revenue <- (N/(N+1)) * 1/((N+1)^(1/N))

revenue <- optimal_price*(maximum_valuation >= optimal_price)
mean(revenue)
expected_revenue
```

What variable captures the number of simulations we are using in the code?

Please enter ONLY the name of the variable without any additional text. Make sure that the capitalization matches the code!

6. Now, perform this exercise for different number of simulations: 10, 100, 1000, and 10000. As you increase the number of simulations, does the mean of the numeric revenue vector coincide more or less than with the analytic solution?

   a) It coincides more

   b) It coincides less

7. Now, we will compare the results we just computed, which hold for the posted price model, with the results we would get from an **auction**. Let's consider an English auction, where buyers optimal strategy is to stay until $p = v_i$ and then leave once $p > v_i$. As shown in lecture, the equilibrium price in this case is the second highest valuation.

   What is the expected revenue when there are two bidders (N=2)? Again, assume that bidders' valuations follow a uniform [0,1] distribution.

   a) For 2 bidders the expected revenue of an English Auction is $\frac{2}{3}$

   b) For 2 bidders the expected revenue of an English Auction is $\frac{1}{3}$

   c) For 2 bidders the expected revenue of an English Auction is 1

   d) For 2 bidders the expected revenue of an English Auction is $\frac{1}{4}$

8. What is the minimum number of bidders such that a buyer prefers to sell the good in an English Auction rather than a posted price auction?

Note: You can do this in two different ways: one is to solve the question mathematically (difficult!), and the other one is to use the simulation in R to answer the question. To use the simulation in R, you will need to write code that computes the expected revenue in an English Auction and the expected revenue in a posted price auction given some number of bidders. You can then compare the two expected revenues for different numbers of bidders.

a) You will need at least 1 bidder.

b) You will need at least 2 bidders.

c) You will need at least 3 bidders.

d) You will need at least 4 bidders.

```
#Preliminaries
rm(list = ls())
library("mvtnorm")
setwd("/Users/raz/Dropbox/14.31 edX Building the Course/Problem
Sets/PSET 5")

#Uniform Valuations
number_of_bidders <- 3
N <- number_of_bidders
V <- 10000
set.seed(5)
valuations <- matrix(runif(
  N*V, min = 0, max = 1),
  nrow = V)

#Posted Price
maximum_valuation <- apply(valuations, 1, max)
optimal_price <- 1/((N+1)^(1/N))
expected_revenue_posted <- (N/(N+1)) * 1/((N+1)^(1/N))
revenue <- optimal_price*(maximum_valuation >= optimal_price)

mean(revenue)
expected_revenue_posted

#Comparison with English Auction
rank_of_valuations <- apply(valuations, 1, rank)
price_english_auction <- apply(valuations, 1, function(x)
(x[rank(x)== N - 1]))
expected_revenue_english <- (N-1)/(N+1)

mean(price_english_auction)
expected_revenue_english
```

9. On the website www.modelingonlineauctions.com, you will find a number of data sets from actual auctions conducted on eBay.

   Download one involving the sale of Cartier watches: *Cartier+3-day+auctions.csv*. There are data on auctions of 18 different watches. For each auction, there is an auction ID, bids, time of each bid, bidder name, bidder rating, minimum bid for the auction, and winning bid for the auction. (Note that the winning bid is not the maximum bid submitted by the highest bidder but rather the second-highest bid plus an increment.)

   Load the data into R. How many auctions are in this data?

Now clean the data set that contains the following variables:

- The id of the auction
- The ratio of the second highest bid to the third highest bid.
- The number of bidders.
- The number of bids.

We can provide you with the following R-code to create these variables, but some information is missing. You will need to either fill the information or create your own code.

```
cartier_data <- read.csv("XXX")
cartier_data$auctionid <- as.character(cartier_data$auctionid)
unique_bids <- unique(cartier_data$XXX)

ratio <- rep(NA, times = XXX)
XXX <- rep(NA, times = length(unique_bids))
number_of_bids <- rep(NA, times = length(unique_bids))

for (i in c(1:length(unique_bids))){
  temp <- subset(XXX, cartier_data$XXX == unique_bids[i])
  bid2 <- XXX[rank(temp$bid, ties.method = 'last') == (length(temp$bid)-1)]
  bid3 <- temp$bid[rank(temp$bid, ties.method = 'last') == (length(XXX)-2)]
  ratio[i] <- bid2 / bid3
  number_of_bidders[XXX] <- length(unique(XXX))
  number_of_bids[i] <- length(temp$bid)
}

data_clean <- data.frame(unique_bids, ratio, number_of_bidders, XXX)
```

10. What is the mean of the ratio of the second highest bid to the third highest bid?

11. What is the median of the number of bidders?

12. What is the maximum value of the number of bids?

13. We can think of ordered bids as being order statistics from some underlying distribution of valuations. Using this perspective, would you expect the number of bidders and the number of bids to inform the ratio between the second and third highest bids?

    (Note: We are not looking for a precise, mathematical answer here, just a bit of informed speculation.)

    a) Yes

    b) No

14. Suppose that the PDF $f_X(x)$ of a random variable X is an even function. ($f_X(x)$ is an even function if $f_X(x) = f_X(-x)$. Is it true that the random variables X and $-X$ are identically distributed?

    a) True

    b) False

15. A couple decides to continue to have children until a daughter is born. What is the expected number of children of this couple if the probability that a daughter is born is given by p?

    a) The expected number of children is given by $\frac{1}{p}$

b) The expected number of children is given by $\frac{1-p}{p}$

c) The expected number of children is given by $\frac{(p}{p^3}$

d) The expected number of children is given by $\frac{1}{p} - 1$

16. Which of the following statements is correct? (Select all that apply)

a) If $f_X(x) = ax^{a-1},\ 0 < x < 1,\ a > 0$ then $\mathbb{E}[X] = \frac{a}{a+1}$

b) If $f_X(x) = \frac{1}{n},\ x = 1, 2, \cdots, n,\ n > 0$ an integer then $\mathbb{E}[X] = \frac{n+2}{2}$

c) If $f_X(x) = \frac{3}{2}(x-1)^2,\ 0 < x < 2$ then $\mathbb{E}[X] = 2$

d) If $f_X(x) = ax^{a-1},\ 0 < x < 1,\ a > 0$ then $\mathbb{E}[X] = 1 + \frac{1}{a}$

e) If $f_X(x) = \frac{3}{2}(x-1)^2,\ 0 < x < 2$ then $\mathbb{E}[X] = 1$

f) If $f_X(x) = \frac{1}{n},\ x = 1, 2, \cdots, n,\ n > 0$ an integer then $\mathbb{E}[X] = \frac{n+1}{n}$

17. Suppose that the random variable Y has a binomial distribution with n trials and success probability X, where n is a given constant and X is a uniform(0,1) random variables. What is $\mathbb{E}Y$?

a) This is given by n

b) This is given by $\frac{n}{2}$

c) This is given by $\frac{n}{3}$

d) This is given by $\frac{X}{n}$

18. Suppose that the random variable Y has a binomial distribution with n trials and success probability X, where n is a given constant and X is a uniform(0,1) random variables. What is $Var(Y)$?

a) This is given by $\frac{n^2}{6} + \frac{n}{12}$

b) This is given by $\frac{n^2}{18} + \frac{n}{6}$

c) This is given by $\frac{n^2}{12} + \frac{n}{6}$

d) This is given by $\frac{n^2}{18} + \frac{n}{12}$

19. Assume that $Y = \alpha + \beta X + U$, where $\beta = \frac{\rho_{XY}\sigma_Y}{\sigma_X}$ and $\alpha = \mu_Y - \beta\mu_X$. What is the expected value of U?

a) The expected value of U is $\alpha + \beta\mu_X$.

b) The expected value of U is μ_Y

c) The expected value of U is 0.

d) The expected value of U is α

20. Assume that $Y = \alpha + \beta X + U$, where $\beta = \frac{\rho_{XY}\sigma_Y}{\sigma_X}$ and $\alpha = \mu_Y - \beta\mu_X$. What is $cov(X, U)$? (Select all that apply)

a) We have that $cov(X, U) = var(X)$

b) We have that $cov(X, U) = \sigma_X\sigma_U$

$c)$ We have that $cov(X, U) = 0$

$d)$ We have that $cov(X, U) = \sigma_X \sigma_Y$

$e)$ We have that $cov(X, U) = \rho_{XU} \sigma_X \sigma_U$