ECO239 Statistics I R Homework 3

DUE: December 30th (Friday) at 12:30
Submit a printed hard-copy.
No Late HWs will be accepted (Even for 5 min). No exception!

- The contents of this homework covers discrete and continuous probability distributions.
- The tasks can be done either by using R console window or by installing "RStudio".
- For reporting, copy & paste figures, codes and output to Word or text document, in addition to your comments.
- This assignment accounts for 10% of your final grade.

*RStudio can be downloaded from https://www.rstudio.com/products/RStudio/

- **Codes for each distribution**:

Distribution		Distribution	
Binomial	*binom*	Uniform	*unif*
Hypergeometric	*hyper*	Normal	*norm*
Poisson	*pois*	Exponential	*exp*

| "d"	returns the height of the probability density function
"p"	returns the cumulative density function

e.g. *dbinom* returns the probability density function for Binomial distribution.
e.g. *pnorm* returns the cumulative density function for Normal distribution.

- **Finding cumulative probabilities**: x_0 is the value of your interest. You should fill in blanks to derive actual probabilities.

Binomial	*pbinom($c(x_0)$, size = , prob= , lower.tail= [TRUE/FALSE])*
Hypergeometric	*phyper($c(x_0)$), m= , n= , k= , lower.tail=[TRUE/FALSE])*
Poisson	*ppois($c(x_0)$), lambda= , lower.tail=[TRUE/FALSE])*
Uniform	*punif($c(x_0)$),min= , max= , lower.tail=[TRUE/FALSE])*
Normal	*pnorm($c(x_0)$),mean= , sd= ,lower.tail=[TRUE/FALSE])*
Exponential	*pexp($c(x_0)$),rate= , lower.tail=[TRUE/FALSE])*

Note: *lower.tail=TRUE* if you want to derive the lower-tail probability.

lower tail   upper tail
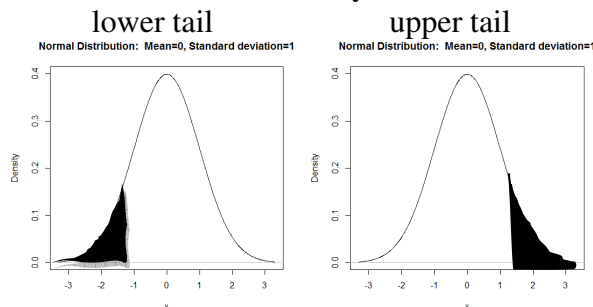
lower.tail=FALSE if you want to derive the upper-tail probability.

- **Plots**

Basic form:

define x values with its minimum, maximum and increment. Set the increment as 0.01.

x<- seq(min, max, increment)

plot(x, dbinom(x, size=n, prob=p),type="h", xlab="Number of Successes", ylab="Probability Mass",main="Binomial Distribution:Trials=n, Probability of success=p")

*Change the entries inside the parenthesis after d_____ (n, p, name of the distribution), labels for x and y axes and main title according to the problem you are working on.

Task1: Discrete Distributions and Probabilities

I: Binomial Distribution

a. Plot probability mass function for Binomial Distribution with the probability of success p = 0.3 and n = 5, 50, 100. When you change p, how does it affect the shapes of the distribution? Comment.
   [Hint: *set x<-seq(0,40,0.01)for all plots*]

b. Create a table for the marginal Binomial probabilities when p = 0.3 and n = 20.
   [Hint: *dbinom(0:n, size=n, prob=p)*]

c. Create a table for the cumulative Binomial probabilities when p = 0.3 and n = 20.
   [Hint: *pbinom(0:n, size=n,prob=p)*]

d. 15% of students admitted to University A will actually enroll.
      i.   What's the probability that more than 8 students will enroll if 30 students were offered admissions?
      ii.   If 70% of students admitted to University will actually enroll, what will be the results for i and ii?
   [Hint: Use cumulative distribution. Pay attention to "less than" and "more than" expressions!]

II: Hypergeometric Distribution

- m = the number of white balls in the urn (= S)
- n = the number of black balls in the urn (= N-S)
- k = the number of balls drawn from the urn (= n)

e. Plot probability mass function for Hypergeometric Distribution with N=50, n = 8, S = 5, 10 and 25. When you change the value of S, how does it affect the shapes of the distribution? Comment.
   [Hint: *set x<-seq(0,10,0.01) for all plots*]

f. A company receives a shipment of 50 items. A random sample of 8 items is selected and the shipment is rejected if more than one of these item is defective.What is the probability of rejecting a shipment containing 15 defectives?

III: Poisson Distribution

g. Plot probability mass function for Poisson Distribution with the expected number of success per time unit = 1, 5, 10.When you change the value of lambda, how does it affect the shapes of the distribution? Comment.
   [Hint: set x<- seq(0,22,0.01) for all plots.]

h. The number of accidents in a production facility has a Poisson distribution with a mean of 1.5 per month.
   i. For a given month what is the probability there will be less than 2 accidents?
   ii. What is the probability there will be more than 18 accidents <u>per year</u>?

Task2: Continuous Distribution and Probabilities

IV: Normal Distribution

i. Plot pdfs for Normal distribution when population mean is 0 and the standard deviation is 1 (= standard normal distribution). Change the value of standard deviation to 10 and observe the changes in the shapes of the distribution. Change the value of mean to 10 (the standard deviation =10) and observe the changes in the shapes of the distribution. Comment.
   [Hint: set x<- seq(-30,30,0.01) for all plots.Use type="l" for the plot.]

j. Calculate the probability $P(x > 72)$ when $x \sim N(60,121)$.

k. It is estimated that the time that a well-known pianist, Fazıl Say, spends on stage at his concerts follows a normal distribution with a mean of 110 minutes and a standard deviation of 20 minutes.

   An audience is attempting to record the entire concert using her cell phone, which has a capacity of recording 120 minutes. What is the probability that this capacity will be insufficient to record the entire concert?

V: Exponential Distribution

l. Plot pdfs for Exponential distribution when the lambda parameter is 0.1. Change the parameter to 0.5 and observe the changes in the shapes. Comment.
   [Hint: set x<-seq(0,30,0.01) for all plots.]

m. Calculate the probability $P(t > 4)$ when $t \sim Exp(lambda = 0.4)$.

n. Delivery trucks arrive independently at PTT distribution center. The mean number of trucks arriving per hour is 10. Given that the truck has just arrived, answer the following questions.What is the probability that the next truck will arrive between 5 and 10 minute?

VI: Normal Distribution with Data Application

For this task, use the data "bdims_partial" uploaded on the course webpage. You will find five variables included in the data set. 1. ID, 2. Age, 3. Weight, 4. Height and 5. Sex (1 being male).

(1) Generate Histogram for Height for Male and Female separately.

(2) Generate Q-Q plot (normal probability plot) for Height for Female and evaluate if the distribution is normal or not. Do the same for Weight for Female.

(3) What is the probability that a randomly chosen young adult female is taller than 180 cm?

(4) What is the probability that a randomly chosen young adult females is less than 65 kg?

Note: Data used for this task is adopted partially from the following site:
http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/normal_distribution/normal_distribution.html