ECO239 Group Assignment 2

Due: Dec.29 (Friday) 2017 at 12:30.

*Late homework (even for 5 min.) will NOT be accepted.

*Submission by e-mail will NOT be accepted.

Submit a printed hard-copy together with (1) Honor code and (2) Contribution Sheet

No Late HWs will be accepted (Even for 5 min). No exception!

- The contents of this homework covers discrete and continuous probability distributions.
- Complete these tasks by using **"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.

- **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.

If you set lower.tail=TRUE, it calculates P(X≤x0). If you set lower.tail=FALSE, it calculates P(X>x0).

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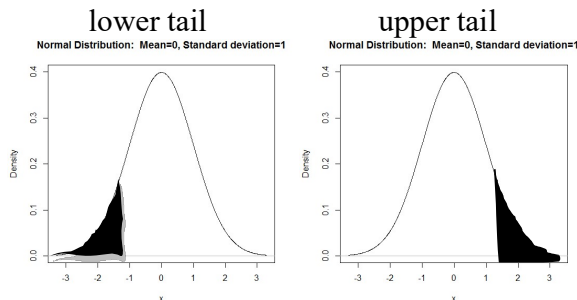
lower.tail=FALSE if you want to derive the upper-tail probability.

- **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. Find the size n which makes probability mass function for Binomial Distribution looks normal (bell-shaped, symmetric) distribution, given p = 0.3. Answer this question by plotting the functions with different n.
   [Hint: *set x<-seq(0,40,0.01)for the plots*]

b. Create a table for the cumulative Binomial probabilities n = 13 and the probability (p) of your choice.
   [Hint: *pbinom(0:n, size=n,prob=p)]*

c. P% of students admitted to University A will actually enroll. Set your own P. What's the probability that more than 20 students will enroll if 30 students were offered admissions?
   [Hint: Use cumulative distribution. Pay attention to "less than" and "more than" expressions!]

II: Poisson Distribution

d. The number of accidents in a production facility has a Poisson distribution with a mean of 2 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 per year?

e. Set up a question by yourself on Poisson Approximation for Binomial Distribution and solve it.

Task2: Normal Distribution

In order to complete this task successfully (and stress-free), refer to the following web page. (Link is also available on the course webpage).

http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/normal_distribution/normal_distribution.html

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) Calculate mean and standard deviation for Weight for Female and Male, separately.

(2) Generate Histogram for Weight for Male and Female separately. Correctly identify the range of X variable.

(3) Generate Boxplot for Weight for Male and Female. Identify the potential outliers.

(4) Generate Normal Q-Q plot (normal probability plot) for Weight for Female and evaluate if the distribution is normal or not. Do the same for Weight for Male. Interpret the distributional shapes based on the generated Normal Q-Q plot.

(5) Take natural log of Weight for Female variable and generate Normal Q-Q plot using the ln(Weight). Report your observation.

(6) What is the probability that the weight of a randomly chosen young adult male is more than 90 kg?

(7) What is the probability that the weight of a randomly chosen young adult male is between 60 and 70 kg?

(8) Write out two probability questions that you would like to answer; one regarding female heights and one regarding female weights and answer the questions.