Lab #7

By the beginning of the next lab (March 18th), send what you have to [afodor@uncc.edu](mailto:afodor@uncc.edu)

Send your code and the answers to questions..

Make sure the text “Lab #7” is in the subject line…

--------------------------------------------

1. Consider a tournament with the following prize structure:

|  |  |
| --- | --- |
| Number of wins | Payoff (in dollars) |
| 0 | 1.45 |
| 1 | 1.72 |
| 2 | 2.24 |
| 3 | 2.76 |
| 4 | 3.55 |
| 5 | 4.6 |
| 6 | 5.65 |
| 7 | 6.75 |

You play in the tournament until you get three losses or 7 wins.

Make a graph of the expected value of the tournament (on the y-axis) vs. the probability of winning each game (on the x-axis). If the tournament costs $3.75, at what win percentage does the expected value exceed the cost of the tournament? (Plot $3.75 as a horizontal line ranging from 0 to 1).

(hint: Calculate the probability of 7 wins a 1 – sum(prob of all other outcomes)).

(2) Generate a simulated dataset in which the variance is a function of the mean but the null hypothesis of differential expression is always true. The following code will accomplish this:

rm(list=ls())

numRows = 3000

numCols = 20

for( i in 1:numCols)

myFrame <- data.frame(1:numRows)

#initiate the data.frame with the correct # of rows to suppress error messages.

#likely, there are much better ways to do this!

names(myFrame)[1] <- "tempColumn"

for( i in 1: numCols)

{

vals <- vector(length=numRows)

for( j in 1:numRows)

{

aMean = j /10

aMean = max( aMean,5)

aVar = aMean+ 5\* aMean

aVal = round( max( rnorm(1,mean=aMean,sd=sqrt(aVar)), 1))

vals[j] = aVal

}

colName <- paste( "sample" , i ,sep="")

myFrame[[colName]] = vals

}

myFrame["tempColumn"] <- NULL

row.names(myFrame) <- paste("Gene\_",1:numRows,sep="")

For each row in the spreadsheet, we will consider the first 10 columns to be “case” and the last 10 to be “control”. For each row in the spreadsheet, run a t-test and generate a p-value for the null hypothesis that the case and control samples follow the same distribution (see slide 5 of lecture 11 for code to do this).

1. Use a simple threshold of p <0.05. How many significant hits would you expect to find if the null hypothesis is always true? How many hits did you actually find in your run?
2. What is the Bonferroni adjusted p-value threshold? What percentage of the time would you expect to see a significant gene under this threshold? How many genes in fact do you see significant at this threshold?

The desired threshold / number of tests

1. Next use a BH FDR corrected threshold of p < 0.05 (which R can do for you with p.adjust(pvals,method="BH") if pvals is a vector that holds the unadjusted p-values). How many hits do you find with the BH adjusted p-values.

* Rank all of the p-values in ascending order. The rank of each value = K
* N \* p / K > threshold

Show your code for all answers