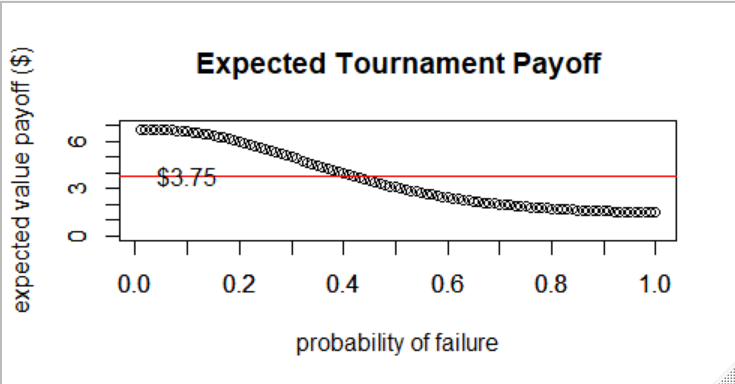
**1. Consider a tournament with the following prize structure:**

|  |  |
| --- | --- |
| 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).**

Graph of the expected value of the tournament (on the y-axis) vs. the probability of losing each game (on the x-axis) is at figure 1. The R code is located at [this link](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab07/generateData.R).



**Figure 1:** Expected Tournament Payoff

**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).**

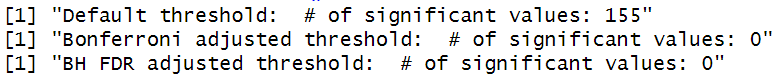
The win percentage of about 60% (i.e., loss percentage of about 40%) exceeds the expected cost of the tournament. See figure 1.

(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.**

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).

Using [this R code](https://github.com/jyoung67/advstatistics-labs/blob/master/labs/lab07/createDataset.R), I generated the following results for A,B & C.

  
**Figure 2:** Various P-value Thresholds

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?**

* I would expect to find 150 (3000\* 0.05) significant hits.
* I observed 155 significant hits (figure 2).

1. **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 Bonferroni adjusted p-value threshold is a calculated value that modifies/adjusts the significance threshold for raw p-values. It is calculated as follows: ***the desired threshold / number of tests***.
* I would expect 0.0017% (0.05/3000) of the time to see a significant gene under this threshold.
* I observed zero significant genes (figure 2).

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.

* I observed zero hits with the BH adjusted p-values:  
  