Statistics 341

Fall 2008 - Assignment #3 Solutions

Due Monday, October 6

This assignment is worth a total of 117 points.

1. Problems from Textbook. See course webpage for problems.

   - (3 pts) Problem 3.1
     Y = number of impurities in well. Y = 0 if neither A nor B occur, $(\overline{A} \cap \overline{B})$; Y = 1 if A occurs and B does not or if B occurs and A does not, $(\overline{A} \cap B) \cup (\overline{B} \cap A)$; Y = 2 if both A and B occur, $(A \cap B)$. The box below gives the given probabilities and the rest of the probability based on these assignments.

	B	$\overline{B}$	
A	0.1	0.3	0.4
$\overline{A}$	0.4	0.2	0.6
	0.5	0.5	1

     So the probability distribution function for Y is

y	0	1	2
p(y)	0.2	0.7	0.1

   - (3 pts) Problem 3.5
     Pictures can be placed only once, and so there are 3! = 6 possible orderings of the three pictures. And since the child is randomly guessing, each of the 6 possible orderings is equally likely. As an example, suppose the three pictures are a Bear, Tiger, and Lion. The child can assign the three pictures as

Bear	Tiger	Lion	3
Bear	Lion	Tiger	1
Tiger	Bear	Lion	1
Tiger	Lion	Bear	0
Lion	Bear	Tiger	0
Lion	Tiger	Bear	1

     This means the random variable Y can have three values, 0, 1 or 3 with probabilities

y	0	1	3
$p(y)$	2/6	3/6	1/6

   - (8 pts) Problem 3.30
     (a) (1 pt) The mean of X will be larger than the mean of Y since each value of X is 1 larger than the corresponding value of Y.
     (b) (2 pts) $E(X) = E(Y + 1) = E(Y) + E(1) = E(Y) + 1$. Yes, this agrees with my result from part (a).
     (c) (2 pts) The variance of X should be equal to the variance of Y. X is defined by just shifting the value of Y by 1. This shifting does not change the spread of the values of X. They will be the same as the spread of the values of Y.

(d) (3 pts)

$$\begin{aligned} E((X - E(X))^2) &= E((Y+1) - (E(Y)+1))^2) \\ &= E((Y - E(Y))^2) \\ &= \sigma^2 \end{aligned}$$

- (8 pts) Problem 3.31
  (a) (1 pt) The mean of W will be larger than the mean of Y since each value of W is 2 times larger than the corresponding value of Y.
  (b) (2 pts) $E(W) = E(2Y) = 2E(Y)$. Yes, this agrees with my result from part (a).
  (c) (2 pts) The variance of W will be larger than the variance of Y. W is defined by multiplying the value of Y by 2. This will increase the spread of the values, resulting in a larger variance.
  (d) (3 pts)

$$\begin{aligned} E((W - E(W))^2) &= E((2Y - E(2Y))^2) \\ &= E(2(Y - E(Y))^2) \\ &= 4E((Y - E(Y))^2) \\ &= 4\sigma^2 \end{aligned}$$

- (8 pts) Problem 3.32
  (a) (1 pt) The mean of U will be larger than the mean of Y since each value of U is 1/10 as large as the corresponding value of Y.
  (b) (2 pts) $E(U) = E(Y/10) = E(Y)/10$ Yes, this agrees with my result from part (a).
  (c) (2 pts) The variance of W will be smaller than the variance of Y. W is defined by multiplying the value of Y by (1/10). This will decrease the spread of the values, resulting in a smaller variance.
  (d) (3 pts)

$$\begin{aligned} E((U - E(U))^2) &= E(((1/10)Y - E((1/10)Y))^2) \\ &= E((1/10)(Y - E(Y))^2) \\ &= (1/10)^2 E((Y - E(Y))^2) \\ &= (1/100)\sigma^2 \end{aligned}$$

- (8 pts) Problem 3.33
  (a) (4 pts)

$$\begin{aligned} E(aY + b) &= E(aY) + E(b) \\ &= aE(Y) + b \\ &= a\mu + b \end{aligned}$$

(b) (4 pts)

$$\begin{aligned} V(aY+b) &= E((aY+b-E(aY+b))^2) \\ &= E((aY+b-(aE(Y)+b))^2) \\ &= E((aY-aE(Y))^2) \\ &= E(a(Y-E(Y))^2) \\ &= a^2E((Y-E(Y))^2) \\ &= a^2\sigma^2 \end{aligned}$$

- (8 pts) Problem 3.34

  In order to calculate the mean and variance of the daily cost, we will need to calculate the mean and variance of the number of times the tool is used Y.

  Y = number of times the tool is used daily.

  $E(Y) = 0(0.1) + 1(0.5) + 2(0.4) = 1.3$

  $E(Y^2) = 0(0.1) + 1(0.5) + 2^2(0.4) = 2.1$

  $V(Y) = E(Y^2) - (E(Y))^2 = 2.1 - 1.3^2 = 0.41.$

  Now define the random variable W = cost of using tool = 10Y

  $E(W) = E(10Y) = 10E(Y) = 10(1.3) = 13.$

  $V(W) = V(10Y) = 100V(Y) = 100(0.41) = 41.$

2. (17 pts) Go to the course webpage and click on the link **Setting Up Random Variables in R**. In this file is code for setting up the sample space obtained when rolling two dice and using this sample space to obtain the random variable for the sum of the two rolls and for the maximum value of the two rolls. In this problem, we will look at a different random variable, the minimum value of the two rolls.

   (a) (5 pts) Use R to set up the sample space and to obtain the random variable, the minimum value of the two rolls. Give the theoretical probability distribution function for the minimum value of the two rolls.

   The sample space can be set up in the same way as in the R help file. Here is the R code.

```
Stwodice<- scan()
1 1 1 2 1 3 1 4 1 5 1 6
2 1 2 2 2 3 2 4 2 5 2 6
3 1 3 2 3 3 3 4 3 5 3 6
4 1 4 2 4 3 4 4 4 5 4 6
5 1 5 2 5 3 5 4 5 5 5 6
6 1 6 2 6 3 6 4 6 5 6 6

Stwodice<- matrix(Stwodice, byrow = T, ncol = 2)
```

   Using the sample space of 36 outcomes, we want to determine the minimum value of the two dice for each outcome. Here is the R code.

```
mintwodice<- apply(Stwodice, 1, min)
```

Getting a table of values of **mintwodice** will give you the number of outcomes out of the 36 that map to each minimum value. This is

```
table(mintwodice)
 1  2  3  4  5  6
11  9  7  5  3  1
```

Since each of the 36 outcomes are equally likely, the probability distribution function for the random variable is

1	2	3	4	5	6
11/36	9/36	7/36	5/36	3/36	1/36

(b) (5 pts) Use R to summarize the theoretical pdf of the random variable with a histogram. (Hint: you need to use appropriate breaks given the possible values for the random variable.) Include the histogram with your assignment.

Here is the R code to produce the histogram. The histogram is located at the end of the assignment.

```
hist(mintwodice, breaks = c(0:6) + 0.5, prob = T)
```

(c) (3 pts) Describe the histogram in part (b).

The histogram is part (b) is skewed to the right. The steps at each minimum value are the same size (2/36).

(d) (4 pts) Use R to find the theoretical mean and standard deviation of the random variable, the minimum value of the two rolls.

Since each of the 36 values in **mintwodice** are equally likely, the mean of this variable will be our theoretical mean value.

```
mean(mintwodice)
[1] 2.527778
```

The standard deviation will be

```
sqrt(var(mintwodice))
[1] 1.424001
```

3. (18 pts) Go to the course webpage and click on the link **An Introduction to the Concept of Random Variables**. In this file is code for rolling two dice and looking at two different random variables from this sample space, the sum of the two rolls and the maximum value of the two rolls. In this problem, we will look at a different random variable, the minimum value of the two rolls.

(a) (5 pts) Use R to roll two dice 10000 times and find the observed probability distribution function for the minimum value of the two rolls. What values will the observed probability distribution function be close to? Why?

Here is the R code to generate the 10000 rolls of the two dice and find the minimum value of the two dice for each of the 10000 rolls.

```
dice<- c(1:6)
dice1<- sample(dice,10000,replace = T)
dice2<- sample(dice,10000,replace = T)
dicematrix<- cbind(dice1,dice2)
mintwodicesim<- apply(dicematrix,1,min)
```

Here are my values for the 10000 rolls for the minimum values of the two dice. My values will be slightly different than your values.

```
table(mintwodicesim)
mintwodicesim
   1    2    3    4    5    6
3007 2561 1924 1343  886  279
```

However, all values should be very similar to the theoretical probability distribution function in part (a) of the problem above. At 10000 rolls, the observed probabilities will be fairly close to the theoretical probabilities.

(b) (5 pts) Use R to make a histogram of the observed probability distribution function. (Hint: you need to use appropriate breaks given the possible values for the random variable.) Include the histogram with your assignment.

Here is the R code to make this histogram. My histogram is included at the end of the assignment.

```
hist(mintwodicesim, breaks = 0:6 + 0.5, prob = T)
```

(c) (3 pts) Describe the histogram in part (b). Your histogram will be similar to the one of the theoretical probability distribution function. Why?

The histogram is part (b) is skewed to the right. The steps at each minimum value are about the same size (2/36). The histogram is not exactly the same as the theoretical probability distribution function since there is some variability in the 10000 observed rolls.

(d) (5 pts) Use R to find the observed mean and standard deviation of the minimum value of the two rolls. How close are the observed mean and standard deviation to the theoretical mean and standard deviation?

Here are my mean and standard deviation. Your values will be similar, but not exactly the same.

```
mean(mintwodicesim)
[1] 2.5377
sqrt(var(mintwodicesim))
[1] 1.408963
```

However, all values should be close to the theoretical mean and standard deviation values. Mine are and yours should be too.

4. (36 pts) The Powerball Lottery has two bins, one containing numbers 1-55, and the other containing numbers 1-42. Five numbers are drawn without replacement from the first bin, and one number, the Powerball, is drawn from the second bin. The cost of playing the Powerball Lottery game is \$1. There are 9 different ways to win when playing Powerball. They are listed in the table below.

Match	Prize	Number of Ways to Match
5 numbers plus Powerball	Grand Prize	1
5 numbers	$200,000	41
4 numbers plus Powerball	$10,000	250
4 numbers	$100	10,250
3 numbers plus Powerball	$100	12,250 (see part (b) below)
3 numbers	$7	502,250 (see part (c) below)
2 numbers plus Powerball	$7	196,000
1 number plus Powerball	$4	1,151,500
Powerball	$3	2,118,760

(a) (2 pts) How many possible sets of 5 numbers plus the powerball exist in the Powerball Lottery game?

From the 55 white balls, you choose 5 and from the 42 powerballs, you choose 1. This makes $\binom{55}{5}\binom{42}{1} = 146107962$

(b) (2 pts) Find the number of ways to match 3 numbers plus the Powerball.

Divide the 55 white balls into the 5 winners and the 50 losers. To match 3 numbers, you have to have 3 numbers from the 5 winners and 2 numbers from the 50 losers. There are $\binom{5}{3}\binom{50}{2} = 12250$ ways to do this. To match the powerball, there is only one way. So we have 12250 ways to match 3 numbers plus the Powerball.

(c) (2 pts) Find the number of ways to match 3 numbers.

The first part is the same. Divide the 55 white balls into the 5 winners and the 50 losers. To match 3 numbers, you have to have 3 numbers from the 5 winners and 2 numbers from the 50 losers. There are $\binom{5}{3}\binom{50}{2} = 12250$ ways to do this. However, this time we cannot match the powerball. There are 41 ways to do this. Making a total of $12250(41) = 502250$ ways to match 3 numbers.

(d) (2 pts) How many possible sets of 5 numbers plus the powerball will win you no prize in the Powerball Lottery game?

You will need to take the answer from part (a) and subtract all the numbers from the table above including the two numbers you calculated in parts (b) and (c). This is

$146107962 - (1 + 41 + 250 + 10250 + 12250 + 502250 + 196000 + 1151500 + 2118760) = 142116660$

(e) (5 pts) Assume the Grand Prize for a particular drawing is \$93.4 million. What is the expected earnings (or loss) when playing the Powerball Lottery for this drawing? (Hint: It costs one dollar to puchase a ticket).

This can be set up in R. The variable winnings is the amount of money you win with each type of ticket.

```
> winnings<- scan()
0 3 4 7 7 100 100 10000 200000 93400000
```

The variable waystomatch are the number of ways to get the corresponding type of ticket.

```
> waystomatch<- scan()
142116660 2118760 1151500 196000 502250 12250 10250 250 41 1
```

The variable probmatch is then the probability of getting the corresponding type of ticket.

```
probmatch<- waystomatch/146107962
```

Once we have this set up, the expected earnings can be calculated using the values of the winnings and their corresponding probabilities. Notice we have subtract 1 from this expectation, since it costs \$1 to purchase a ticket.

```
sum(winnings*probmatch) - 1
-0.1636320
```

(f) (5 pts) Write an equation that gives you the expected earnings (or loss) when playing the Powerball Lottery for a given Grand Prize amount.

Let GP equal the grand prize amount. We can calculated the expected winnings using the equation below.

$$\begin{aligned} E(\text{win}) &= (GP(1) + 200000(41) + 10000(250) + 100(10250) + 100(12250) + \\ &\quad 7(502250) + 7(196000) + 4(1151500) + 3(2118760) + 0(142166660))/(146107962) \\ &= (GP + 28800030)/146107962 \end{aligned}$$

The expected earnings is found by subtracting 1 from the expected winnings.

$$E(\text{earnings}) = (GP + 28800030)/146107962 - 1 = (GP - 117307932)/146107962$$

(g) (5 pts) Enter your equation into R, and calculate the expected earnings (or loss) when playing the Powerball Lottery for Grand Prizes of 50, 100, 125, 150, 200, 250, 300, 350 and 400 million dollars.

You can use R to calculate the expected earnings for the different values of the grade prize. The variable gp sets up the grand prize amounts.

```
gp<- scan()
50000000 100000000 125000000 150000000 200000000
250000000 300000000 350000000 400000000
```

The variable below is the expected earnings function.

```
expectedearnings<- (gp-117307932)/146107962
> expectedearnings
[1] -0.46067258 -0.11845988  0.05264647  0.22375282  0.56596552  0.90817821
[7]  1.25039091  1.59260361  1.93481631
```

(h) (5 pts) Using R, obtain a plot of the Grand Prize amount vs. expected earnings (or loss).

To plot these values, use the R command

```
plot(gp, expectedearnings)
```

The plot is at the end of this assignment.

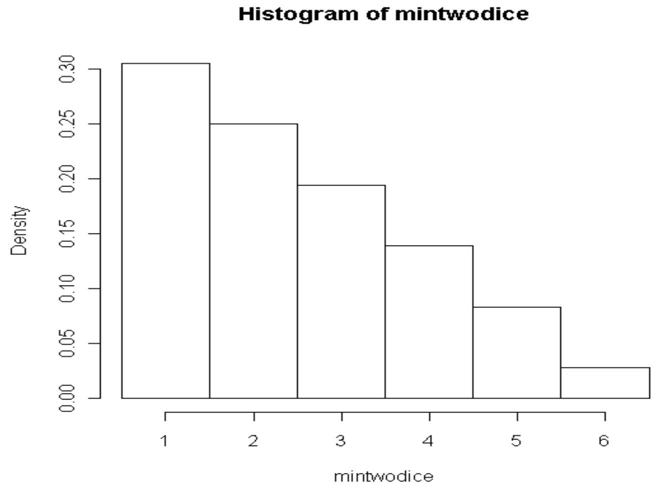
(i) (4 pts) How much would the Grand Prize for a particular drawing have to be in order for the expected earnings to be \$0? Calculate this value and draw this value on your plot.

As you can see from the values above and from the plot, this grand prize value is somewhere between 100 and 125 million dollars. The exact value is

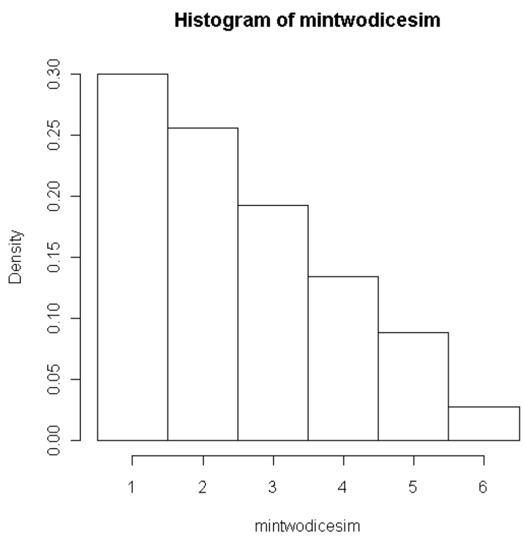
$$(GP - 117307932)/146107962 = 0 \qquad \text{or} \qquad GP = 117307932$$

(j) (4 pts) In doing these calculations, we have ignored at least one factor in determining how much money you will actually receive from any of these prizes. What have we ignored? How will this factor change the calculations above?

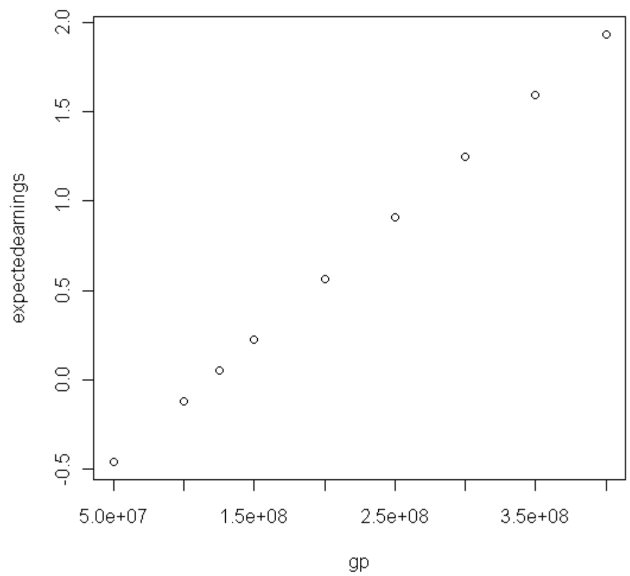
We have actually ignored three factors. One is how many people win the grand prize. If more than one person wins the grand prize, the amount is split equally between the winners. The second thing is the way in which you accept the grand prize, either as one lump sum payment or as an annuity. And the third thing we have ignored is taxes on the amount you win. All of these things will change the calculations above, decreasing either the actual grand prize amount or decreasing each winning amount (except for probably the very small winnings) due to taxes.



Above is the histogram of the theoretical probability distribution function for the minimum value of the two dice.



Above is the histogram of the observed probability distribution function for the minimum value of the two dice.



Above is the plot of the grand prize amounts vs. the expected earnings for the Powerball Lottery.