Homework 2

Problem 1. Dice Rolls

If you roll a pair of fair dice, what is the probability of..

(a) getting a sum of 1?

 0 cells below sum to 0

(b) getting a sum of 5?

  4 cells below sums to 5   4/36 = 0.1111 alternatively using R

  > getEventProb(1,2,6,list(5),FALSE)

  [1] 0.1111111

(c) getting a sum of 12?

  1 cell below sums to 12   1/36 = 0.027777 alternatively using R:

  > getEventProb(1,2,6,list(12),FALSE)

  [1] 0.02777778

	1	2	3	4	5	6
1	1,1	2,1	3,1	4,1	5,1	6,1
2	1,2	2,2	3,2	4,2	5,2	6,2
3	1,3	2,3	3,3	4,3	5,3	6,3
4	1,4	2,4	3,4	4,4	5,4	6,4
5	1,5	2,5	3,6	4,5	5,5	6,5
6	1,6	2,6	3,6	4,6	5,6	6,6

Problem 2. School absences

Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.

(a) What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?

P(0 absence) 1-0.68 = 0.32 0.32

(b) What is the probability that a student chosen at random misses no more than one day?

P(no more than 1 absence) = (0.32 + 0.25) = 0.57

(c) What is the probability that a student chosen at random misses at least one day?

P(misses 1 or more) = (0.25 + 0.15 + 0.28) = 0.68

(d) If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.

Assume the kids missing school or not are independent events.

P(kid 1 not missing) * P(kid 2 not missing) = 0.32* 0.32 = 0.1024

(e) If a parent has two kids at a DeKalb County elementary school, what is the probability that both kids will miss some school, i.e. at least one day? Note any assumption you make.

Assume the kids missing school or not are independent events.

P(at least 1 missing) * P(at least 1 missing) = 0.68* 0.68 = 0.4624

(f) If you made an assumption in part (d) or (e), do you think it was reasonable? If you didn't make any assumptions, double check your earlier answers.

Perhaps not reasonable, because if one kid gets sick, the chances of the other getting sick may actually increase, thus violating my independent events assumption.

0	0.32
1	0.25
2	0.15
3	0.28
Total	1.00

Problem 3. Health coverage, relative frequencies

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table displays the distribution of health status of respondents to this survey (excellent, very good, good, fair, poor) and whether or not they have health insurance.
mat=matrix(c(.023, 0.0364, 0.0427, 0.0192, 0.0050,0.2099,

```
0.3123 ,0.2410 ,0.0817,0.0289), byrow=TRUE, nrow=2)
colnames(mat)=c("Excellent", "Very Good","Good", "Fair","Poor")
rownames(mat)=c("No Coverage","Coverage")
mat
##          Excellent Very Good  Good   Fair   Poor
## No Coverage   0.0230   0.0364 0.0427 0.0192 0.0050
## Coverage      0.2099   0.3123 0.2410 0.0817 0.0289
```

(a) Are being in excellent health and having health coverage mutually exclusive?

No, 0.2099 are in excellent health and have coverage.

(b) What is the probability that a randomly chosen individual has excellent health?

P(Excellent health) = (0.0230 + 0.2099) = 0.2329

(c) What is the probability that a randomly chosen individual has excellent health given that he has health coverage?

P(Excellent | Coverage) = (2099 / 8738) = 0.2402

(d) What is the probability that a randomly chosen individual has excellent health given that he doesn't have health coverage?

P(Excellent | No Coverage) = (230/ 1262) = 0.1823

(e) Do having excellent health and having health coverage appear to be independent?

No, P(excellent health) * P(has coverage) != P(excellent health and has coverage)
 (0.2329 * 0.8733) = 0.2034 != 0.2099

Problem 4. Exit Poll.

Edison Research gathered exit poll results from several sources for the Wisconsin recall election of Scott Walker. They found that 53% of the respondents voted in favor of Scott Walker. Additionally, they estimated that of those who did vote in favor for Scott Walker, 37% had a college degree, while 44% of those who voted against Scott Walker had a college degree. Suppose we randomly sampled a person who participated in the exit poll and found that he had a college degree. What is the probability that he voted in favor of Scott Walker?

Given:
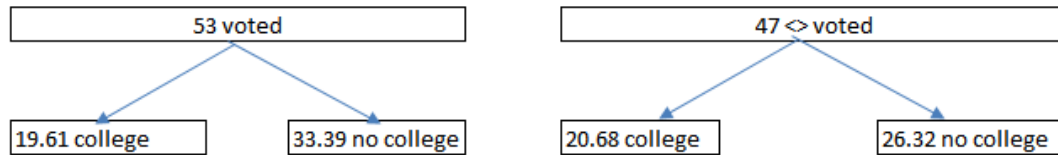

```
> prob <- 19.61 / (19.61 + 20.68)
> prob
[1] 0.4867213
```

Problem 5. Books on a bookshelf

The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

mymat2=matrix(c(13,59,15,8),nrow=2,byrow=TRUE)
colnames(mymat2)=c("hard","paper")
rownames(mymat2)=c("fiction","nonfiction")

mymat2
hard paper
fiction   13   59
nonfiction   15   8

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

```
> a <- ((13 +15)/ (13+15+59+8)) *(59/(13+15-1+59+8))
> a
[1] 0.1849944
```

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.

```
> b <- ((13+59)/(13+15+59+8)) * ((12+15)/(12+59+15+8))
> b
[1] 0.2176932
```

(c) Calculate the probability of the scenario in part (b), except this time complete the calculations under the scenario where the first book is placed back on the bookcase before randomly drawing the second book.

```
> c <- ((13+59)/(13+15+59+8)) * ((13+15)/(13+59+15+8))
> c
```

[1] 0.2233795

(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

Because (27/94) ~ (28 /95)

Problem 6. Is it worth it?

Andy is always looking for ways to make money fast. Lately, he has been trying to make money by gambling. Here is the game he is considering playing: The game costs 2 dollars to play. He draws a card from a deck. If he gets a number card (2-10), he wins nothing. For any face card (jack, queen or king), he wins 3 dollars. For any ace, he wins 5 dollars and he wins an extra $20 if he draws the ace of clubs.

(a) Create a probability model and find Andy's expected profit per game.

```
Expected_return <- (36/52)*0 + (12/52)*3 + (4/52)*5 + (1/52)* 20
> Expected _return
[1] 1.461538
```

(b) Would you recommend this game to Andy as a good way to make money? Explain.

No, his expected return is $1.46 on a $2.00 investment - a loss of $0.54.

Problem 7. Scooping ice cream.

Ice cream usually comes in 1.5 quart boxes (48 fluid ounces), and ice cream scoops hold about 2 ounces. However, there is some variability in the amount of ice cream in a box as well as the amount of ice cream scooped out. We represent the amount of ice cream in the box as X and the amount scooped out as Y . Suppose these random variables have the following means, standard deviations, and variances:

```
mymat3=matrix(c(48,1,1, 2,.25,.0625), nrow=2, byrow=TRUE)
colnames(mymat3)=c("mean", "SD", "Var")
rownames(mymat3)=c("X, In Box","Y, Scooped")
mymat3
##            mean  SD   Var
## X, In Box   48   1.00  1.0000
## Y, Scooped  2   0.25  0.0625
```

(a) An entire box of ice cream, plus 3 scoops from a second box is served at a party.

How much ice cream do you expect to have been served at this party? What is the standard deviation of the amount of ice cream served?

```
> ice_cream <- 48 +(2*3)
> ice_cream
[1] 54


> sqrt_icecream <- sqrt(1+0.0625*3)
> sqrt_icecream
[1] 1.089725 = standard deviation
```

(b) How much ice cream would you expect to be left in the box after scooping out one scoop of ice cream? That is, find the expected value of $X - Y$. What is the standard deviation of the amount left in the box?

```
> expectvalueX <- 48 -2
> expectvalueX
[1] 46

> standarddev <- sqrt(1 + 0.0625)
> standarddev
[1] 1.030776
```

(c) Using the context of this exercise, explain why we add variances when we subtract one random variable from another.

I'm not sure I follow this question.