STAT 515 hw 3

Random variables, expected value, hypergeometric, binomial

1. Let X be the sum of two dice rolls.

   (a) Complete the table by finding the probabilities $P(X = x)$ and $P(X \leq x)$ for $x = 2, \ldots, 12$.

x	2 3 4 5 6 7 8 9 10 11 12
$P(X = x)$	
$P(X \leq x)$	

   The probability distribution of X is given by

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
$P(X \leq x)$	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36

   (b) Find $P(X \geq 10)$ and interpret this value in words.

   It is $6/36 = 0.167$, so in every 36 rolls, we expect to see 6 for which the sum is greater than or equal to 10.

   (c) In the board game Settlers of Catan, players roll two dice at the beginning of each turn. The player moves "the robber" if the sum of the rolls is 7. On what proportion of turns do you expect this to occur?

   If many many turns were taken, the expected proportion of turns on which the robber is moved is equal to $6/36 = 1/6 = 0.167$.

   (d) Find $\mathbb{E}X$.

   We obtain
   $$\mathbb{E}X = \frac{2 \cdot 1 + 3 \cdot 2 + \cdots + 12 \cdot 1}{36} = \frac{252}{36} = 7.$$

   (e) Find $\text{Var}\, X$.

   $$\text{Var}\, X = \frac{(2-7)^2 \cdot 1 + (3-7)^2 \cdot 2 + \cdots + (12-7)^2 \cdot 1}{36} = \frac{210}{36} = 5.833333.$$

   (f) Suppose you rolled two dice 1,000 times, recording each time the sum of the two rolls; afterwards you compute the average of these 1,000 numbers. To what value would you expect this average to be close?

> We would expect this number to be close to the expected value of X, which is 7.

(g) Suppose you rolled two dice 1,000 times. Each time you take the sum of the rolls, subtract 7, square the result, and record this number. Afterwards you compute the average of these 1,000 numbers. To what value would you expect this average to be close?

> We would expect this number to be close to the variance of X, which is 5.833333.

(h) Check your answers to the previous two parts by running a simulation in R: As a starting point, use the command

$$\texttt{R1 <- sample(c(1:6),1000,replace=TRUE)}$$

to simulate rolling a 6-sided die 1,000 times. Now figure out how to simulate rolling two dice 1,000 times, recording each time the sum of the two rolls. Then compute the mean and variance of these roll sums.

> We can use the commands
>
> ```
> R1 <- sample(c(1:6),1000,replace=TRUE)
> R2 <- sample(c(1:6),1000,replace=TRUE)
> X <- R1 + R2
> mean(X)
> mean((X - 7)^2)
> ```
>
> I got
>
> ```
> > mean(X)
> [1] 7.056
> > mean((X - 7)^2)
> [1] 5.488
> ```

2. Use the survey data from the first day of class (download from course website) to answer this question. Let X be the number of siblings of a randomly selected student from our class.

(a) Tabulate the probability distribution of X.

> From the data we have
>
> | x | 0 | 1 | 2 | 3 | 4 |
> |---|---|---|---|---|---|
> | $P(X = x)$ | 2/38 | 20/38 | 14/38 | 1/38 | 1/38 |

(b) Compute the expected value of X.

This is
$$0(2/38) + 1(20/38) + 2(14/38) + 3(1/38) + 4(1/38) = 1.447368$$

(c) Compute the average of the values in the `sibs` column of the data set.

This is also 1.447368.

(d) The median of a random variable X is the smallest value x such that $P(X \leq x) \geq 1/2$. Give the median of X.

The median of X is 1.

3. Suppose a coin is flipped four times and let Y be the number of heads.

(a) Give the sample space S for the experiment of flipping a coin 4 times.

The sample space is
$$S = \left\{ \begin{array}{llll} HHHH & HHHT & HHTT & \\ & HHTH & HTTH & \\ & HTHH & TTHH & TTTH \\ & THHH & THHT & TTHT \\ & & THTH & THTT \\ & & HTHT & HTTT & TTTT \end{array} \right\}$$

(b) Write down the probability distribution of Y in tabular form

$$\begin{array}{c|c} y & \\ \hline P(Y = y) & \end{array}$$

We have

y	0	1	2	3	4
$P(Y = y)$	1/16	4/16	6/16	4/16	1/16
$P(Y \leq y)$	1/16	5/16	11/16	15/16	16/16

(c) What is the expected value of Y? Show calculations.

$\mathbb{E}Y = 2$.

(d) What is the variance of Y? Show calculations.

> $\operatorname{Var} Y = 1$

(e) Regarding Y as a random variable with a binomial distribution, what are n and p for the binomial distribution?

> They are $n = 4$ and $p = 1/2$.

(f) Compute np.

> We have $np = 4(1/2) = 2$.

(g) Compute $np(1 - p)$.

> We have $4(1/2)(1 - 1/2) = 1$.

4. In a shipment of 100 parts, 4 are defective. The purchaser has time to test only 10 randomly selected parts for defects. Let X be the number of defective parts the purchaser finds.

(a) Give the support $\mathcal{X}$ of X.

> The support is $\mathcal{X} = \{0, 1, 2, 3, 4\}$.

(b) What is the probability that the purchaser finds no defective parts?

> We have
> $$P(X = 0) = \frac{\binom{4}{0}\binom{96}{10}}{\binom{100}{10}} = 0.6516305.$$

(c) What is the probability that the purchaser finds at least 1 defective part?

> We have
> $$P(X \leq 1) = 1 - P(X = 0) = 1 - 0.6516305 = 0.3483695.$$

(d) Suppose the shipment has d defective parts. What is the largest value of d such that the probability is at least 0.90 that the purchaser finds no defects among 10 randomly sampled parts?

> We find that with $d = 1$ defective part, the probability that the purchaser finds no defects in 10 randomly sampled parts is $\binom{1}{0}\binom{99}{10}/\binom{100}{10} = 0.90$. With $d = 2$ defective parts, the probability is $\binom{2}{0}\binom{98}{10}/\binom{100}{10} = 0.8090909$. Since the probability continues to get smaller as d increases, the answer is $d = 1$.

5. Suppose that in a class of 20 people, 6 of them plan to attend graduate school. Suppose you take a sample of 5 students from the class *without replacement*, letting X be the number of students in your sample who plan to attend graduate school.

   (a) Tabulate the distribution of X in the format

$$\begin{array}{c|c} x & \\ \hline P(X = x) & \end{array}$$

---

x	0	1	2	3	4	5
$P(X = x)$	0.129	0.387	0.352	0.117	0.014	0.000

We can get these probabilities from R using the command

```
round(dhyper(x = 0:5, m = 6, n = 20 - 6 , k = 5),3)
```

or by putting x <- 0:5 and running

```
round(choose(6,x)*choose(20-6,5-x) / choose(20,5),3)
```

   (b) Suppose that you draw a single student from the class and *then replace them*, doing this 5 times and letting X be the total number of times that the student drawn planned to attend graduate school. Tabulate the distribution of X in the format

$$\begin{array}{c|c} x & \\ \hline P(X = x) & \end{array}$$

---

x	0	1	2	3	4	5
$P(X = x)$	0.168	0.360	0.309	0.132	0.028	0.002

We can get these probabilities from R using the command

```
round(dbinom(x = 0:5, size = 5 , prob = 6/20),3)
```

   (c) Consider a class of 200 students of which 60 plan to attend graduate school. Suppose you took a sample of 5 students from this class *without replacement*, letting X be the number of students in your sample who plan to attend graduate school. Tabulate the distribution of X in the format

$$\begin{array}{c|c} x & \\ \hline P(X = x) & \end{array}$$

x	0	1	2	3	4	5
$P(X = x)$	0.164	0.363	0.312	0.131	0.027	0.002

We can get these probabilities from R using the command

```
round(dhyper(x = 0:5, m = 60, n = 200 - 60, k = 5),3)
```

or by putting x <- 0:5 and running

```
round(choose(60,x)*choose(200-60,5-x) / choose(200,5),3)
```

(d) Under what circumstance is the difference between sampling with replacement and sampling without replacement negligible?

When the population is large, sampling with replacement and sampling without replacement will give very similar results, since the composition of a large population changes very little when a member is sampled and not replaced.

6. Suppose you take a random sample of 30 USC undergraduates and ask them if they recycle their aluminum cans.

(a) If 9,932 of the 24,830 USC undergraduates recycle their aluminum cans, what is the exact probability (according to the hypergeometric distribution) that less than half of the 30 students in the sample recycle their aluminum cans?

We can compute this using

```
phyper(q = 14, m = 9932, n = 24830 - 9932, k = 30)
```
$= 0.824773$

of by putting x <- 0:14 and executing

```
sum(choose(9932,x)*choose(24830-9932,30-x)/choose(24830,30))
```
$= 0.824773.$

(b) For a random variable X having the binomial distribution with $n = 30$ and $p = 0.4$, what is $P(X \leq 14)$?

```
round(pbinom(q = 14, size = 30, prob = .40),4)
```
$= 0.8246.$

(c) Why can the number of students X who recycle their aluminum cans in a random sample of 30 USC undergraduates be regarded as a random variable with the binomial distribution?

The population is so big that we can treat sampling without replacement like sampling with replacement.

(d) What is the expected value of X?

$$\mathbb{E}X = (0.4)30 = 12.$$

(e) What is the variance of X?

$$\text{Var}\, X = 30(0.4)(0.6) = 7.2$$