

Statistics 324 Homework #3

*Submit your homework to Canvas by the due date and time. Email your lecturer if you have extenuating circumstances and need to request an extension.

*If an exercise asks you to use R, include a copy of the code and output. Please edit your code and output to be only the relevant portions.

*If a problem does not specify how to compute the answer, you may use any appropriate method. I may ask you to use R or use manual calculations on your exams, so practice accordingly.

*You must include an explanation and/or intermediate calculations for an exercise to be complete.

*Be sure to submit the HWK3 Auto grade Quiz which will give you ~20 of your 40 accuracy points.

*50 points total: 40 points accuracy, and 10 points completion

Probability

Exercise 1 You are adding Badger-themed bedazzle to your striped overalls and are using both red and white beads. You are interested in how the size of the bag of beads you select your beads from changes the probability of outcomes of interest. Compute the probability for outcomes a and b using three different sampling strategies each time.

(Small Pop) drawing without replacement from a small population where the bag of beads contains 6 White beads and 4 Red beads.

(Large Pop) drawing without replacement from a large population where the bag of beads contains 600 White beads and 400 Red beads.

(Same Pop) drawing from a population where the bag of beads always contains 60% White and 40% Red beads.

Example: Consider choosing 3 beads. Calculate the probability of selecting no white beads.

Small Pop: $P([RRR]) = \frac{4}{10} * \frac{3}{9} * \frac{2}{8} = 0.03333333$

Large Pop: $P([RRR]) = \frac{400}{1000} * \frac{399}{999} * \frac{398}{998} = 0.06371181$

Same Pop: $P([RRR]) = 0.40 * .40 * .40 = 0.064$

a. Consider choosing 3 beads. Calculate the probability of selecting exactly 1 white bead.

Small Pop:

```
((6/10) * (4/9) * (3/8)) + ((4/10) * (6/9) * (3/8)) + ((4/10) * (3/9) * (6/8))
```

```
## [1] 0.3
```

Large Pop:

```
((600/1000) * (400/999) * (399/998)) + ((400/1000) * (600/999) * (399/998)) + ((400/1000) * (399/999) *
```

```
## [1] 0.2881439
```

Same Pop:

```
(6/10 * 4/10 * 4/10) * 3
```

```
## [1] 0.288
```

b. Consider choosing 3 beads. Calculate the probability of selecting at least 1 white bead.

Small Pop:

```
1 - ((4/10) * (3/9) * (2/8))
```

```
## [1] 0.9666667
```

Large Pop:

```
1 - ((400/1000) * (399/999) * (398/998))
```

```
## [1] 0.9362882
```

Same Pop:

```
1 - ((4/10) * (4/10) * (4/10))
```

```
## [1] 0.936
```

c. Consider sampling *without replacement*. Does drawing from a population that is **small** or **large** relative to your sample size result in a probability that is closest to the probability when sampling *with replacement*?

Drawing from a **large** population relative to your sample size results in a probability that is closest to the probability when sampling with replacement.

Exercise 2 Six hundred paving stones were examined for cracking and discoloration. Eighteen (18) were found to be cracked and 24 were found to be discolored. A total of 562 stones were neither cracked nor discolored.

a. Create a 2-way table to organize the counts of stones for each of the 4 combinations of Cracked/Not Cracked and Discolored/Not Discolored.

X	Cracked	Not Cracked	Total
Discolored	4	20	24
Not Discolored	14	562	576
Total	18	582	600

- b. What is the probability that a randomly sampled paving stone from this set is discolored and not cracked?

20/600

[1] 0.03333333

- c. What is the probability that a randomly sampled paving stone from this set is cracked or discolored?

38/600

[1] 0.06333333

- d. What is the probability that in a random sample of 3 paving stones from this set (without replacement), at least one of the three is cracked or discolored?

1 - ((562/600) * (561/599) * (560/598))

[1] 0.1784993

- e. What is the probability that a randomly sampled paving stone from this group has discoloration, given we know that it is cracked?

4/18

[1] 0.2222222

- f. Is being discolored and cracked independent in this set of 600 paving stones?

No, it is not independent, since the probability of a stone being discolored changes if we know if the stone is cracked as well.

- g. Now suppose in another group of 600 paving stones, forty-eight (48) were found to be cracked and 25 were found to be discolored. How many stones would be cracked and discolored if the events: discolored and cracked are independent in this group of 600 stones? Make sure to show how you calculated your answer.

X	Cracked	Not Cracked	Total
Discolored	??		25
Not Discolored			575
Total	48	552	600

I knew the number of cracked and discolored stones had to be less than 25, since that was the total number of discolored stones.

Exercise 3 Four patients make appointments to have their blood pressure checked at a clinic. Let X be the number of them who have high blood pressure. Based on data from the National Health and Examination Survey, the approximate probability distribution of X based on long term data for this type of patient is:

X	0	1	2	3	4
P(X=x)	0.22	0.40	0.28	0.09	0.01

- a. What is the value of $P(X \geq 2)$? What does this value mean in the context of the question?

0.38. This means that the probability that more than 2 of the four people selected have high blood pressure is 38%.

- b. What is the probability that at least one of the 4 patients will have high blood pressure?

78%.

- c. Compute the expected value of X , μ_X . What does this value mean in the context of the question?

```
(0 * 0.22) + (1 * 0.40) + (2 * 0.28) + (3 * 0.09) + (4 * 0.01)
```

```
## [1] 1.27
```

This means that the mean amount of people with high blood pressure is about 1.27. d. What is the standard deviation of X , σ_X ? What does this value mean?

```
sqrt(0.22 * (0 - 1.27) ** 2 + 0.40 * (1 - 1.27) ** 2 + 0.28 * (2 - 1.27) ** 2 + 0.09 * (3 - 1.27) ** 2 + 0.01 * (4 - 1.27) ** 2)
```

```
## [1] 0.9365362
```

This value means that it is much more likely for the number of people in the range of $[1.27 - 0.936, 1.27 + 0.936]$ to have high blood pressure than people outside this range.

- e. Consider using a binomial random variable with $n=4$ to approximate the distribution of X given above, $X \sim \text{Bin}(n = 4, \pi = ??)$. What is an approximate probability of a single patient of this type having high blood pressure when they make an appointment, π ?

$$\pi = 1.27/4 = \boxed{0.3175}$$

```
dbinom(4, 4, 0.3175)
```

```
## [1] 0.0101619
```

Exercise 4 A distributor receives a very large shipment. The distributor would like to accept the shipment if 15% or fewer of the items are defective and to return it if more than 15% of the components are defective.

Someone on the quality assurance team samples 4 items. Let X be the random variable for the number of defective items in the sample. You can assume that the defectiveness of items is independent within the shipment. Also assume that exactly 15% of the items in the shipment are defective.

- a. Determine the probability distribution of X (write out the pmf) using probability theory.

x	$P(X=x)$
0	0.5220062
1	0.368475
2	0.0975375
3	0.011475
4	0.00050625

```
# Parameters
n <- 4 # Number of items sampled
p <- 0.15 # Probability of an item being defective

# Calculate the PMF for X
pmf <- dbinom(0:n, size = n, prob = p)

# Print the PMF
for (k in 0:n) {
  cat("P(X = ", k, " ) = ", pmf[k + 1], "\n")
}
```

```
## P(X = 0 ) = 0.5220062
## P(X = 1 ) = 0.368475
## P(X = 2 ) = 0.0975375
## P(X = 3 ) = 0.011475
## P(X = 4 ) = 0.00050625
```

- b. Compute the probability that $X > 0$. What does this value mean in the context of the scenario?

$$1 - 0.5220062 = \boxed{0.4779938}$$

- c. Compute the expected value for X , μ_X using a *short* computation. What does value mean in the context of the scenario?

$$4 * 0.15 = \boxed{0.6}$$

- d. Compute the standard deviation for X , σ_X using a *short* computation.

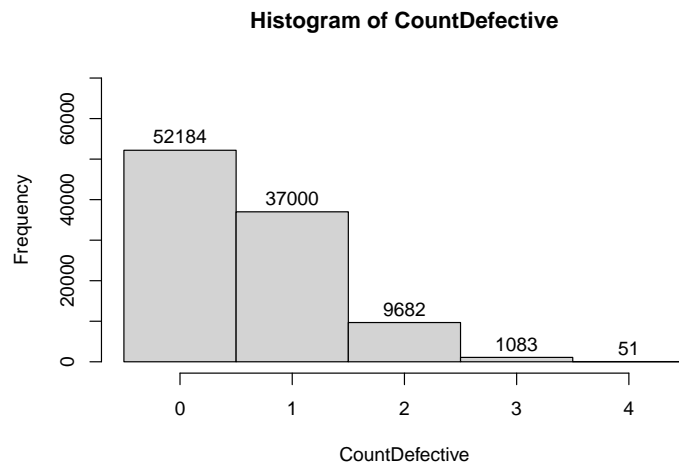
$$\sqrt{4 * 0.15 * (1 - 0.15)} = \boxed{0.7141428}$$

- e. Explain (briefly ~ 2 sentences) how you can use the following simulation to check your answers for part a. Some questions to consider: Why did I define `IsDefective` as I did? What values are stored into the `CountDefective` vector? What does the histogram show?

You can use the following simulation to check your answers for part a because it makes a Monte Carlo simulation to measure the probability of selecting a defective item. IsDefective makes a vector of 1s and 0s to simulate the probability of selecting a defective item (a 1) as opposed to a non-defective item. The CountDefective vector stores the results of running the simulation many times, and is converted to a histogram to show the amount of defects in each simulation.

```
IsDefective=c(rep(1,15), rep(0,85))
manytimes=100000
CountDefective=rep(0,manytimes)
set.seed(1)
for (i in 1:manytimes){
  samp=sample(IsDefective,4, replace=TRUE)
  CountDefective[i]=sum(samp)
}

hist(CountDefective, labels=TRUE,
     ylim=c(0,.7*manytimes), breaks=seq(-0.5, 4.5, 1))
```



f. Suppose the quality assurance employee is now going to look at 20 items from the shipment and believes it is reasonable to use a Binomial model $X \sim \text{Bin}(n = 20, \pi = 0.15)$ to describe the number of items in those 20 that will have a defect.

fi. Compute the probability that exactly 5 of those 20 items have a defect.

```
dbinom(5, 20, 0.15)
```

```
## [1] 0.1028452
```

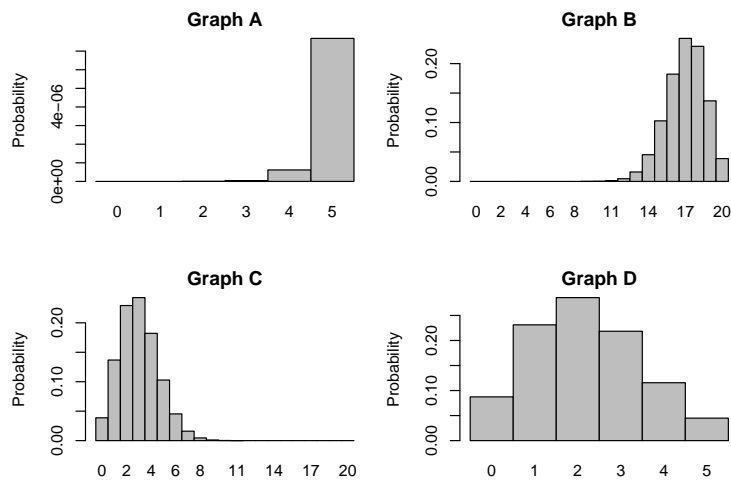
fii. Compute the probability that 5 or more of those 20 items have a defect.

```
1 - pbinom(4,20,0.15)
```

```
## [1] 0.1701532
```

fiii. Which histogram given below correctly shows the probability distribution for the binomial model described in f? How do you know?

```
par(mfrow=c(2,2), mar=c(4,4,2,1))
barplot(names=0:5,dbinom(0:5, 15, prob=0.85),
        xlab="", ylab="Probability", main="Graph A", space=0)
barplot(names=0:20, dbinom(0:20, 20, prob=0.85),
        xlab="", ylab="Probability", main="Graph B", space=0)
barplot(names=0:20, dbinom(0:20, 20, prob=0.15),
        xlab="", ylab="Probability", main="Graph C", space=0)
barplot(names=0:5, dbinom(0:5, 15, prob=0.15),
        xlab="", ylab="Probability", main="Graph D", space=0)
```



```
par(mfrow=c(1,1), mar=c(5.1, 4.1, 4.1, 2.1))
```

Graph C shows the correct probability distribution for the binomial model described in f. It has the correct probability of 0.15, as well as being right-skewed which makes sense – the higher the value, the lower the probability should be in this case.