Spring 2021 Midterm II Gradescope Portion SOLUTIONS

Your name here

Today's date

By typing your name above, you indicate that you had no prior knowledge of the content of the exam, did not share the exam with anyone, and did not access any unauthorized resources while taking the exam including but not limited to other people and internet resources not explicitly provided by this course.

Question 1 [3 points total]

Cystic fibrosis (CF) is caused by a gene mutation that affects the cells that produce mucus, sweat, and digestive juices. The mucus becomes thick and sticky, causing severe damage to the respiratory, digestive and reproductive systems. Children who inherit only one copy of the defective gene are carriers who could pass the gene on to their own children. But children who inherit two copies of the defective gene—one copy from each parent—will develop cystic fibrosis. A study in Northern Europe tested 543 children for the defective gene and also for CF infection. In all, 25% of children had at least one copy of the defective gene, and 6.6% of the children had both the defective gene and CF. For the following questions, drawing a Venn diagram may be helpful.

1.1 [1 point] What is the probability that a child has CF given that the child has the defective gene?

```
# SOLUTION: P(CF|DG) = P(CF & DG) / P(DG) = 0.066 / 0.25 = 0.264
```

1.2 [1 point] What is the probability that a child has CF given that the child does not have the defective gene?

```
# SOLUTION:

#P(No DG) = 1 - 0.25 = 0.75

#P(CF & No DG) = 0.346 - 0.066 = 0.28

#P(CF|No DG) = 0.28/0.75 = 0.373
```

1.3 [1 point] What is the relationship between CF and the defective gene? Select all that apply.

(a) Disjoint
(b) Independent
(c) Dependent
(d) Mutually Exclusive

```
# SOLUTION: (c) Dependent
```

Question 2 [2 points total]

Having unattached earlobes is a dominant trait, while having attached earlobes is recessive. Bob and his wife are expecting to have a child. Bob is heterozygous for his unattached earlobes, while his wife has attached earlobes.

2.1 [1 point] Write the sample space for their child's earlobe type in the correct notation.

```
# SOLUTION: S = {Attached earlobes, unattached earlobes}
```

2.2 [1 point] The sample space is...

(a) discrete
(b) continuous
(c) neither

```
# SOLUTION: (a) discrete
```

Question 3 [1 point total]

3.1 [1 point] Which parameters define the normal distribution? Select all that apply.

(a) mode
(b) mean
(c) standard deviation
(d) amount of data

```
# SOLUTION: (b) and (c).The mean defines the position of the normal distribution
# on the x-axis.The standard deviation defines the height and spread of the normal distribution.
```

Question 4 [1 point total]

In a group of 800 athletes, 5 in 8 athletes have never smoked e-cigarettes regularly. You choose 3 of the athletes at random and ask about their e-cigarette use status. The distribution of the number you choose who have never smoked e-cigarettes regularly is:

A _____ distribution with the parameters n = _____ and p = _____.

First blank:

Second blank:

Third blank:

```
# SOLUTION: binomial, 800, 0.625
```

Question 5 [1 point total]

A researcher has a data set with a mean score of 45 and standard deviation of 5. Assuming that the data set is normally distributed, the researcher would expect approximately 95% of the data to fall between what values?

(a) 20 and 70
(b) 35 and 55
(c) 40 and 50
(d) 30 and 60

```
# SOLUTION: (b) 35 and 55
```

Question 6 [2 points total]

During a four-year study period, there were 5,368 distinct surgical hospitalizations in which at least one adverse event was reported after surgery, an average rate of 10 adverse events reported for every 154 surgical hospitalizations (6.5%). Assume surgical hospitalizations are independent of each other and the rate of adverse events is constant. Let X be the count of adverse events after 154 surgical hospitalizations.

6.1 [1 point] Write a line of code using a function learned in lab to find the probability that no adverse events will occur after 154 surgical hospitalizations.

```
# SOLUTION: dpois(x = 0, lambda = 10)
```

6.2. [1 point] You want to use R to calculate the probability that more than 7 adverse events occur after 154 surgical hospitalizations. Which of the following lines of code will NOT give you the correct output?

(a) `1 - ppois(q = 7, lambda = 10)`
(b) `ppois(q = 7, lambda = 10, lower.tail = F)`
(c) `1 - (dpois(x = 7, lambda = 10) + dpois(x = 6, lambda = 10) + dpois(x = 5, lambda = 10) + dpois(x = 4, lambda = 10) + dpois(x = 3, lambda = 10) + dpois(x = 2, lambda = 10) + dpois(x = 1, lambda = 10) + dpois(x = 0, lambda = 10))`
(d) `1 - (dpois(x = 6, lambda = 10) + dpois(x = 5, lambda = 10) + dpois(x = 4, lambda = 10) + dpois(x = 3, lambda = 10) + dpois(x = 2, lambda = 10) + dpois(x = 1, lambda = 10) + dpois(x = 0, lambda = 10))`

```
# SOLUTION: (d)
```

Question 7 [1 point total]

7.1 [1 point] Select all of the *false* statements about the central limit theorem.

(a) When the sample size is large, the sampling distribution is approximately Normal, no matter what the underlying distribution looked like.
(b) The larger the sample, the better the approximation.
(c) The standard deviation of the sampling distribution gets smaller as n increases.
(d) As the sample size increases, the sampling distribution becomes more centered around the median of the true population.

```
# SOLUTION: (b) and (d)
```

Question 8 [1 point total]

8.1 [1 point] Assuming an unbiased sample, the confidence level is the long term success rate of the method that produces the interval.

(a) True
(b) False

```
# SOLUTION: (a) True
```

Question 9 [1 point total]

9.1 [1 point] You read that a statistical test at the alpha = 0.01 level has probability 0.14 of making a Type II error when a specific alternative is true. What is the power of the test against this alternative?

(a) 0.76
(b) 0.28
(c) 0.86
(d) 0.99

```
# SOLUTION: (c) 0.86
```

Question 10 [1 point total]

Read the following abstract from a study titled: The probability of a cancer cluster due to chance alone.

We propose to use a very simple model to test whether a cancer cluster is due to chance alone. We focus on the acute childhood leukaemia cluster in Columbus, Ohio. In 1975, 12 leukaemia cases were observed in Columbus while the expected number is 6 cases per year. According to our simple model, the probability of such an occurrence, due to chance alone, is less than 1 per cent. However, if we divide the child population of the U.S.A. into 200 regions (each region having 200,000 children) then the probability that at least one region will see, in a given year, 12 or more cases is higher than 80 per cent. So in this sense the Columbus cluster could be attributed to chance alone. However, the probability that any of the 200 regions see 18 cases or more in a given year is almost 0. Thus, a cluster of 18 or more cases in a region of 200,000 children in one year should be regarded as highly suspicious and should be investigated.

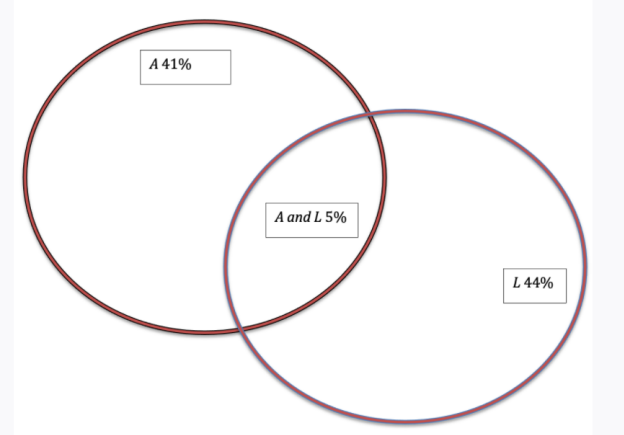
10.1 [1 point] Which type of distribution would be appropriate for this data?

(a) Normal
(b) Poisson
(c) Binomial
(d) None of the above

```
# SOLUTION: (b) Poisson
```

Question 11 [1 point total]

Thrombophlebitis is due to one or more blood clots in a vein that cause inflammation. The probability that an individual has blood clots in their arms is P(A)=0.41. The probability that an individual has blood clots in their legs is P(L)=0.44. The probability that an individual has blood clots in their arms AND their legs is P(A and L)=0.05. What is the probability that an individual has blood clots in their arms OR in their legs?



END