Research Questions using Lung Cancer Dataset

2.1.)

Suppose an adult has 2 symptoms of being a smoker, and we are interested in seeing if these 2 symptoms are an indicator of cancer. Let X denote that an adult has chest pain and Y that the adult coughs a lot. Let X Y denote that the adult has more severe chest pain and only mild coughing. There are four points in the set S of possible observations:

Let A denote the subset of possibilities containing no coughing; B, the subset containing severe and mild coughing; and C, the subset containing at least one sign of coughing. List the elements of A, B, C, A ∩ B, A ∪ B, A ∩ C, A ∪ C, B ∩ C, B ∪ C, and C ∩ B.

**Answer**:

A = {XX}

B = {YY}

C = {XY, YX, XX}

A ∩ B =

*B*∩*C* = {YY}

*C* ∩ = {YX, XY}

*A*∪*B =* {XX,YY}

*=* S

*=*{XY,YX,XX}

2.14)

A survey classified a large number of adults according to whether they were smoking and if they had lung cancer. The proportions falling into the four resulting categories are given in the following table:

Lung Cancer

|  |  |  |
| --- | --- | --- |
| Smoking | Yes | No |
| yes | .52 | .05 |
| No | .34 | .09 |

If a single adult is selected from the large group, find the probabilities of the events defined below. The adult

**a**The person smokes

**b**The person smokes but does not have lung cancer

**c**The person has lung cancer

**Answer**:

1. .52 + .05 = .57
2. .05
3. .52 + .34 = .86

2.28)

There are four common symptoms that lung cancer patients can possess and these patients usually only have ½ of the symptoms. One and only one patient have a very rare symptom. The positions are filled by choosing two of the applicants at random.

**a**List the possible outcomes for this experiment.

**b**Assign reasonable probabilities to the sample points.

**c**Find the probability that the applicant from the minority group is selected for a position.

**Answer**:

1. {
2. 1/6
3. = .5

2.35)

An adult with lung cancer has 10 doctors to choose from for his treatment. The adult also has 5 treatments to choose from but can only pick one. If the treatment are to be made on by choosing one doctor and one treatment, how many different options does the adult have to get his cancer treated?

**Answer**: 50 different types of options

2.72)

For a certain population of adults, the percentage having or not having a lung cancer, listed according to sex, were as shown in the accompanying table. That is, of all the people taking the exam, 46% were in the male-cancer category, .05% were in the male-no cancer category, and so forth. An employee is to be selected randomly from this population. Let A be the event that the adult has cancer and let M be the event that a male is chosen.

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome | Male | Female | Total |
| Cancer () | 145 | 125 | 270 |
| No Cancer | 17 | 22 | 39 |
| Total | 162 | 147 | 309 |

**A** Are the events A and M independent?

**b** Are the events A and F independent?

**Answer**:

Note that P(A) = .46 and P(A|M) = 39/309 = .12, so yes they are independent. Similarly, P(|F) = 22/309 = .07=P(), so A and F are independent

2.94)

A smoker does two tests, A and B, to check if they have lung cancer. If a symptom is present, the probability that it will be detected by device A is .95; by device B, .90; and by both devices, .88.

**a**If lung cancer is present, find the probability that the smoke will be detected by either device A or B or both devices.

**b**Find the probability that lung cancer will be undetected.

**Answer**:

2.124)

In a survey conducted among a group of adults, it was found that 87% of them have cancer and 13% don’t have cancer. Among those who have cancer, 46% are male and among those who don’t have cancer, 56% are female. If a student is chosen at random and is found to be female, what is the conditional probability that she doesn’t have cancer?

**Answer:**

The conditional probability that she doesn’t have cancer is 12.9%

3.38)

A doctor of a clinic wishes to compare the taste appeal of a new medicine (solution B) with that of the medicine(solution A). Each of 5 patients is given three glasses in random order, two containing solution A and the other containing solution B. Each patient is asked to state which glass he or she most enjoyed. Suppose that the two solutions are identical. Let Y be the number of judges stating a preference for the new solution.

**a**Find the probability function for .

**b**What is the probability that at least three of the five patients state a preference for the new solution?

**c**Find the expected value of

**d**Find the variance of .

**Answer:**

1. P(y) = , y = 0, 1, 2, 3, 4, 5
2. P(Y.5 = 50%
3. E(Y) = 5(1/3) = 5/3
4. V(Y) = 5(1/3)(2/3) = 10/9

3.102)

A hospital has 8 lung cancer patients, 4 breast cancer patients, and 6 pancreatic cancer patients. A patient is chosen from the hospital, one at a time without replacement to get financial assistance with medical bills. What is the probability that all three patients will have lung cancer.

**Answer:**

3.103)

A dataset contains 309 adults, 172 which have allergies. A doctor selects five of the adults at random, thinking all don’t have an allergy. What is the probability that all five of the adults don’t have an allergy?

**Answer:**

Hypergeometric distribution

3.122)

Patients arrive at a checkin counter in a clinic according to a Poisson distribution at an average of 6 per hour. During a given hour, what are the probabilities that

**a**no more than three patients arrive?

**b**at least two patients arrive?

**c**exactly five patients arrive?

**Answer:**

Let Y = # of patients that arrive during the hour. Then, is Poisson with .

4.13)

A cancer treatment center has a 150-gallon tank of medicine that is filled at the beginning of each week. The weekly demand shows a relative frequency behavior that increases steadily up to 100 gallons and then levels off between 100 and 150 gallons. If denotes weekly demand in hundreds of gallons, the relative frequency of demand can be modeled by

**a**Find .

**b**Find

**c**

**Answer:**

1. = F(.5) = 1.2 – ½ - 1/8

Lung Cancer Survey Questions and Word Problems

To write these questions, I used a lung cancer dataset that I obtained from [Kaggle](https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer). In my questions, I tried my best to incorporate the things we have learned in class to get a better understanding of what this data means, such as Poisson distribution, Hypergeometric distribution, and conditional probability. With this dataset and using the information I learned from answering these questions, I concluded that males seemed to be more susceptible in getting lung cancer from females. In question 2.72, I found that males were 6% more likely to get lung cancer than their female counterparts. Moving on, I found that to relate my database to the questions in the textbook, I had to develop scenarios that these adults with cancer were in. This is what lead to questions like 3.38 and 2.35, and I was still able to build situations that seemed relevant to the overall dataset. Although, as I progressed in chapters, I found relating to the database proved to be more and more difficult with the introduction to integrals, so I limited the questions I derived from the book between chapters two to four.

I think my favorite question I developed from using this data is 2.124. This question related to finding the conditional probability that a woman from this dataset doesn’t have cancer. I found this quite interesting because there was only a 12.9% chance that the woman didn’t have cancer. To me, this meant that this dataset had an overwhelming amount of people with lung cancer. This also led me to believe that this is because the creators of this dataset wished to find indicators of lung cancer in adults since there was an abundant amount of people with cancer in this dataset. Although, this makes me wonder why didn’t they only survey people with lung cancer and not include people who don’t have it, since this may lead to varying results.

In conclusion, using this lung cancer dataset granted me the opportunity to develop interesting word problems that were able to relate to the statistical concepts we went over in class. Throughout this process, I was able to draw conclusions and stumble upon newfound information that I wouldn’t have learned just by looking at this dataset. I was also able to successfully create realistic scenarios that were translated into word problems and improved my understanding of statistics. Overall, this project has given me the chance to apply the things I learned in class to real-life scenarios and draw interesting conclusions of lung cancer that I didn’t have before.