Part 2 – An Introduction to Conditional Probability

As you were able to see by playing Pig, the fact that the probability in a given situation can change greatly affects how a situation is approached and interpreted. This sort of idea is prevalent across society, not just in games of chance. Knowledge of **conditional probability** can inform us about how one event or factor affects another. Say-No-To-Smoking campaigns are vigilant in educating the public about the adverse health effects of smoking cigarettes. This motivation to educate the public has its beginnings in data analysis. Below is a table that represents a sampling of 500 people. Distinctions are made on whether or not a person is a smoker and whether or not they have ever developed lung cancer. Each number in the table represents the number of people that satisfy the conditions named in its row and column.

	Has been a smoker for 10+ years	Has not been a smoker
Has not developed lung cancer	202	270
Has developed lung cancer	23	5

1. How does the table indicate that there is a connection between smoking and lung cancer?

2. Using the 500 data points from the table, you can make reasonable estimates about the population at large by using probability. 500 data values is considered, statistically, to be large enough to draw conclusions about a much larger population. In order to investigate the table using probability, use the following outcomes:

   S – The event that a person is a smoker
   L – The event that a person develops lung cancer

   Find each of these probabilities (write as percentages):

   a) $P(S)$
   b) $P(\overline{S})$
   c) $P(L)$
   d) $P(\overline{L})$
   e) $P(L \cap S)$
   f) $P(\overline{S} \cap \overline{L})$
   g) $P(\overline{S} \cap L)$
   h) $P(S \cap \overline{L})$
   i) $P(S \cup L)$
   j) $P(\overline{S} \cup \overline{L})$

3. In order to use probability to reinforce the connection between smoking and lung cancer, you will use calculations of *conditional probability.*

a) By considering only those people who have been smokers, what is the probability of developing lung cancer?

b) Compare the value in 3a to the one for *P(L)* in 2c. What does this indicate?

c) You should be able to confirm that a non-smoker is less likely to develop lung cancer. By considering only non-smokers, what is the probability of developing lung cancer?

4. When calculating conditional probability, it is common to use the term "given." In question 3a, you have calculated the probability of a person developing lung cancer given that they are a smoker. The condition (or, "given") is denoted with a single, vertical bar separating the probability needed from the condition. The probability of a person developing lung cancer given that they are a smoker is written $P(L \mid S)$.

a) Rewrite the question from 3c using the word "given."

b) Write the question from 3c using set notation.

5. Find the probability that a person was a smoker given that they have developed lung cancer and represent it with proper notation.

6. Find the probability that a given cancer-free person was not a smoker and represent it with proper notation.

7. How does the probability in number 6 compare to $P(\overline{L} \mid \overline{S})$? Are they the same or different and how so?

8. Based upon finding the conditional probabilities make an argument that supports the connection between smoking and lung cancer.