Class 6: Conditional Probability (Text: Sections 4.5)

Ex. Use the fact that 13% of the men and 13% of women are left-handed to fill in the following table. What proportion of people are left-handed?

	Men	Women	Total
Left-handed			
Right-handed			
Total	6000	4000	10,000

The probability of a person (either man or woman) being left-handed is 1300/10,000 = 0.13

How this table relates to Independence of handedness and gender:
The probability of a man being left-handed is
Probability of a woman being left-handed is
Probability of a person being left-handed is
Since all three probabilities are

Ex. Use the fact that 7% of men and 0.4% of women are color blind to fill in the following table. What proportion of people are color blind?

	Men	Women	Total
Color blind			
Not CB			
Total	6000	4000	10,000

How this table relates to Independence of colorblindness and gender:
The probability of a man being color blind is
Probability of a woman being color blind is
Probability of a person being color blind is 0

Since the probabilities are

Conditional Probability
A **conditional probability** tells us the likelihood of one event occurring *given that* another event has occurred.

We write conditional probabilities with a vertical line which is read as "given that" or "conditional on", so

P(Color Blind | Male) =

Similarly, we have

P(Color Blind | Female) =

Note: The ordinary probability $P(\text{Color blind}) = 4.36\%$ can be called an *unconditional probability* if we want to distinguish it from a conditional probability.

To calculate conditional probability:

For ***any*** events A and B (not just independent events), we have

$$P(A|B) = \frac{\text{Number of occurences of } A \text{ and } B}{\text{Number of occurences of } B}$$

This can be written as

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Ex: What is P(Male|Color Blind)? What does it tell us? Explain why it has the value it does.

Ex: Find P(Female|Color Blind). Explain why the number you get has the magnitude it does.

Ex: Describe in words the difference between P(Color Blind|Male) and P(Male|Color Blind)

Testing for Independence: We can use any one of the following

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

Or $P(A|B) = P(A)$

Or $P(B|A) = P(B)$

In general, for any events (not just independent):

$P(A \text{ and } B) = P(A|B) \cdot P(B)$ and $P(A \text{ and } B) = P(B|A) \cdot P(A)$

<u>Example: Smoking and Lung Disease</u>

According to the American Lung Association, 7% of the US population has lung disease. Of those with the disease, 90% are smokers; of those without the disease, 25.3% are smokers.[1]

<u>Ex</u>: With no calculation, decide if smoking and lung disease are independent.

<u>Ex</u>: Use the data given to fill in the joint probability distribution table below.

	Lung Disease	No Lung Disease	Total
Smoking			
Non Smoking			
Total			

<u>Ex</u>: What is the probability that a randomly selected smoker has lung disease?

<u>Ex</u>: What is the probability that a randomly selected non-smoker has lung disease?

<u>Ex</u>: Why do the answers to last two questions not add to 1?

[1] www.lung.org. Reported in *Intro Statistics* 9th ed, p. 193, by N. Weiss (Pearson, 2012)

<u>Testing for Independence using conditional probabilities</u>
<u>Ex</u>: Using smokers and non-smokers, calculate conditional probabilities to check that lung disease and smoking are not independent.

<u>Testing for Independence using multiplication of probabilities</u>
<u>Ex</u>: Use the probability that someone has lung disease and is a smoker to show that lung disease and smoking are not independent:

<u>Example: Medical Testing: HIV Screening</u>
If we have a medical screening test, for example a mammogram or an HIV test, several outcomes are possible:

- A true positive (test is positive and person has the disease)
  A false positive
- A true negative
- A false negative

How likely are each type? What does this depend on?

The **prevalence** of the disease is the rate at which the disease occurs in the population. For the US:

P(HIV) =

Accuracy of a medical test
The **sensitivity** of the test is the likelihood that test results are positive among patients with the disease. For current HIV tests:

Sensitivity = P(Test+ | HIV) =

The **specificity** of the test is the likelihood that the test results are negative among patients without the disease. Currently, for HIV

Specificity = P(Test- | No HIV) =

Unfortunately, as the sensitivity increases (the test recognize smaller traces), the specificity tends to decrease.

<u>Ex</u>: In practice we want to know the probability of having HIV if we have a positive test result. What conditional probability do we want to know? This is called the **positive predictive value**.

<u>Ex</u>: Fill in the table to calculate the probability of having HIV if given a positive test result.

For a Sample of 10,000 People

	HIV	No HIV	Total
Positive test			
Negative test			
Total			

Notice that we have to fill in the marginal values first and then the center of the table.

<u>Ex</u>: Since the test is so accurate, explain using a Venn diagram why the positive predictive value is only 60.1%.