Day 7 Worksheet

Activity 1: Risk Assessment

In the book *Gut Feelings*, the author describes a study where he asked 24 doctors to estimate the following probability. Only 8 of the doctors were anywhere close!
G. Gigerenzer Gut Feelings: The Intelligence of the Unconscious, Penguin Books 2008

"In your clinic the probability that one woman has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram. Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?" Make your best guess for the probability:

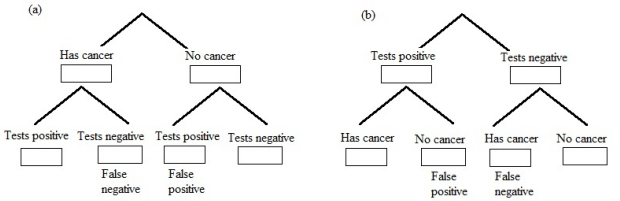
(a) Less than 25% (b) Between 25% and 50%
(c) Between 50 and 75% (d) More than 75%

Justify your thinking to your neigbour.

Activity 1, continued: Assume that 100,000 women come through the clinic. Fill in the contingency table. Hint: In this example it makes sense to fill in the bottom row first.

	Has cancer	No cancer	Total
Tests positive			
Tests negative			
Total			

A flowchart or tree can also be used to organize your thinking. For the problem described above, why is flowchart (a) a better idea than flowchart (b) for this particular problem?



1. What proportion of women with cancer get a positive result?
   $P(positive \mid cancer) =$
2. What proportion of women without cancer get a positive result?
   $P(positive \mid cancer) =$
3. If a woman gets a positive result, is it more likely that she is from the cancer subset of the population or the non-cancer subset?
4. If a woman gets a positive test result, what is the likelihood she has cancer?
   $P(cancer \mid positive) =$
5. If a woman gets a positive test result, what is the chance she does not have cancer?
   $P(no\ cancer|positive) =$
6. What portion of tests results are correct?
   $P(true\ negatives + true\ positives) =$
7. If a woman gets a negative result, what is the chance she does not have cancer?
   $P(no\ cancer \mid negative) =$
8. Why would doctors want to keep this false negative number as low as possible? i.e. Why do we want it to be so much lower than the other boxes?

Activity 2: Let's return to the *videoGamesXBoxPS3* data set from Moodle. Recreate your pivot table in Excel and answer the conditional probability questions below.

	M for mature	T for teen	E+10	E for everyone	Total
PS3					
Xbox					
Total					

For each of the following, be careful in your selection of the divisor. Is the divisor the full count of video games, or a specific subset?

1. If a game is selected at random, what's the chance it is rated M?
2. If a game is rated M, what's the chance it's for PS3?
3. If a game is rated M, what's the chance it's for XBox?
4. If a game is for Xbox, what's the likelihood it is rated E or E+10?
5. If a game is for PS3, what's the likelihood it is rated E or E+10?
6. Problems (2) and (3) should sum to 1. Why?

Problems (4) and (5) likely do not sum to 1. Why?

Activity 3: A certain disease has a prevalence of 2% in a population. Let's assume that a test for the disease is developed and it is estimated that the test is 99% accurate. In this case, "accurate" means that if you have the disease the test is positive and if you do not have the disease the test is negative.

Use your intuition to estimate the probability that a person with a positive test actually has the disease.

Assume there are 10,000 people in the population. Determine the probability via the contingency table.

	Has Disease	No Disease	Total
Tests positive			
Tests negative			
Total			

Activity 3, continued:
Now use a flowchart to organize the probability.

Find the following probabilities

1. What is the probability of a true positive test? A true negative test?
2. What is the probability of a false positive test? A false negative test?
3. What is the probability of a negative test given that you have the disease?
4. What is the probability of having the disease if you test negative?

Activity 4: Binge drinking and car accidents
For men, binge drinking is defined as having five or more drinks in a row, and for women as having four or more drinks in a row. According to a study by the Harvard School of Public Health, 44% of college students engage in binge drinking, 37% drink moderately, and 19% abstain entirely. Another study, published in the American Journal of Health Behavior, finds that among binge drinkers aged 21-34, 17% have been involved in an alcohol-related automobile accident, while among non-binge drinkers of the same age, only 9% have been involved in alcohol-related accidents.

Create a contingency table to organize the information provided in the paragraph. Note that some labels are provided in the table, but you will need to choose the other labels before filling in the table.

	binge	non-binge	total
total			

Activity 4, continued. Once the table is complete, find the following probabilities.

1. Probability that a randomly selected student is a binge drinker
2. Probability that a randomly selected binge drinker has been in an accident
3. Probability that a randomly selected student is not a binge drinker (either drinks moderately or abstains)
4. Probability that a student who drinks moderately or abstains has been in an accident
5. Probability that someone who has been in an accident is a binge drinker
6. Probability that someone who has been in an accident is not a binge drinker