Drill set 1

1. Calculate the probability of flipping a balanced coin four times and getting each pattern: HTTH, HHHH and TTHH.

HTTH = (½) \* (½) \* (½) \* (½) = 1/16 = .0625

HHHH = (½) \* (½) \* (½) \* (½) = 1/16 = .0625

TTHH = (½) \* (½) \* (½) \* (½) = 1/16 = .0625

1. If a list of people has 24 women and 21 men, then the probability of choosing a man from the list is 21/45. What is the probability of not choosing a man?

P(not\_man) = P(woman) = 24/45

1. The probability that Bernice will travel by plane sometime in the next year is 10%. The probability of a plane crash at any time is .005%. What is the probability that Bernice will be in a plane crash sometime in the next year?

P(Bernice in a plane crash) = P(Bernice in a plane) \* P(Plane crash) = .10 \* .00005 =

0.0005%

1. A data scientist wants to study the behavior of users on the company website. Each time a user clicks on a link on the website, there is a 5% chance that the user will be asked to complete a short survey about their behavior on the website. The data scientist uses the survey data to conclude that, on average, users spend 15 minutes surfing the company website before moving on to other things. What is wrong with this conclusion?

The most obvious thing to me is that you might have people who very frequently click on links on the site. These people will likely be included in the survey data. People who do not click on links will likely not be included in the survey data. Seems likely that the survey data will be biased towards the behavior of the people who frequently click on links.

Drill set 2

A diagnostic test has a 98% probability of giving a positive result when applied to a person suffering from Thripshaw's Disease, and 10% probability of giving a (false) positive when applied to a non-sufferer. It is estimated that 0.5 % of the population are sufferers. Suppose that the test is now administered to a person whose disease status is unknown. Calculate the probability that the test will:

1. Be positive

P(+ test) = P(Diseased)\*P(Diseased testing +) + P(Non-diseased)\*P(Non-diseased testing+)

P(+ test) = 0.0005\*.98 + 0.9995\*.10 = 0.00049 + .09995 = 0.10044

1. Correctly diagnose a sufferer of Thripshaw's

.98

1. Correctly identify a non-sufferer of Thripshaw's

.90

1. Misclassify the person

Probability of a + test when person is not-diseased plus the probability of a - test when the person is diseased.

.10\*.9995 + .02\*.0005 = .0996