Probabilty and Biology

Stat 145

Bob's Beautiful Boxes

This doesn't have much to do with biology. But the methods used here will help us with the biological examples that follow.

Bob has two beautiful boxes of balls. Box A contains 2 green balls and 7 red balls. Box B contains 4 green balls and 3 red balls. Bob flips a coin to randomly select a box. Then he randomly selects one ball from that box. If the selected ball is red, what is the probability that it was chosen from Box A?

Let's build up to this step by step.

a. These balls are not equally likely to be selected. Why not? Which are more likely and which less likely?

   Hint: Imagine you get to put a sticker on one ball. If that ball is selected, you win a prize. Assuming you want the prize, where should you put the sticker?

b. Let's define some events.

   - A: Bob selects Box A
   - B: Bob selects Box B
   - R: Bob selects a red ball

c. Let's do some inventory. Which of these probabilities can we easily determine? Which is our main question? How can you use the items on the list that you know to figure out the values you don't know? (Feel free to add additional probabilities to the inventory if that is helpful.)

$$P(A) \qquad P(B) \qquad P(R)$$

$$P(A \,\text{and}\, B) \qquad P(A \,\text{or}\, B) \qquad P(A \,\text{and}\, R) \qquad P(B \,\text{and}\, R)$$

$$P(A \mid R) \qquad P(R \mid A) \qquad P(B \mid R) \qquad P(R \mid B)$$

Breast cancer screening

Here is some information about breast cancer screening. (Note: These percentages are approximate, and very difficult to estimate.)

- American Cancer Society estimates that about 1.7% of women have breast cancer. http://www.cancer.org/cancer/cancerbasics/cancer-prevalence

- Susan G. Komen For The Cure Foundation states that mammography correctly identifies about 78% of women who truly have breast cancer. http://ww5.komen.org/BreastCancer/AccuracyofMammograms.html

- An article published in 2003 suggests that up to 10% of all mammograms are false positive. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1360940

If a mammogram yields a positive result, what is the probability that patient has cancer?

Disease Testing

A test for a medical condition that affects 1 person in 1,000 has the following properties: If a person is healthy, it correctly diagnoses this 98% of the time. If a person is diseased, it correctly diagnoses this 99% of the time. If you take the test and it comes back positive (ie, the test says you have the disease), what is the probability that you have the disease?

Hint: Use the inventory method. Useful events include

- H: person is healthy
- D: person is diseased
- P or $+$: test is positive (indicates disease)
- N or $-$: test is negative (indicates healthy)

Plants

1. Suppose that the red flower-color allele is dominant and is denoted as R, and that the white flower-color allele is recessive and is denoted as r. You have one parent with genotype RR and the red phenotype. The other parent has genotype Rr, and the red phenotype. What is the probability that an offspring plant will be white?

2. Now suppose that you have two parents, each with the Rr genotype. What proportion of offspring will have the red phenotype in this situation?

3. Now suppose we have a plant where R and r are codominant. The Rr genotype results in a pink flower-color. If each of the parent plants is pink, what is the probability that the offspring will be pink? white? red?

People

4. In humans, there is a gene locus for blood type. The three alleles are A, B, and O. A and B are codominant to each other and dominant to O. Genotypes AA and AO give blood type A. Genotypes BB and BO give blood type B. Genotype AB gives blood type AB. And genotype OO gives blood type O. A man of genotype AB marries a woman of genotype BO. What is the probabilty that their kids will have each blood type?

5. A woman with blood type B marries a man with blood type O. If they have a child of blood type O, what is the probability that the mother's genotype is BO? If they have a child of blood type B, can you compute the probability that the mother's genotype is BO; that the mother's genotype is BB? Why?/why not?

6. In the US, the allele frequencies for blood type are approximately 67% O, 25% A, and 8% B. What are the proportions of people with blood types O, A, B, and AB (assuming mating is independent of blood type)?

7. Do you think mating is independent of blood type? Why might it be or not be independent?

8. Cystic fibrosis is a disease caused by a recessive allele. (You need two copies to get cystic fibrosis. If you have only one copy, you are called a carrier.) Across the caucasian population, 2.5% of the alleles are the recessive, disease causing mutation. If mating is independent of cystic fibrosis genetics,

   a. What proportion of the caucasion population will have cystic fibrosis?
   b. What proportion of the caucasion population will be carriers for cystic fibrosis?
   c. What proportion of people without cystic fibrosis are carriers?
   d. If both parents are known to be carriers, what is the probability that their children will have cystic fibrosis?
   e. Why might mating not be independent of cystic fibrosis genetics?

More Medical Tests

Calculating probabilities for diagnostic tests is done so often in medicine and public health that the topic has some specialized terminology.

- The **sensitivity** of a test is the probability of a positive test result when disease is present.
- The **specificity** of a test is the probability of a negative test result when disease is absent.
- The probability of disease in a population is referred to as the **prevalence**.
- The **positive predictive value** (PPV) is the probability that disease is present when a test result is positive.
- The **negative predictive value** (NPV) is the probability that disease is absent when test results are negative.

1. Using the following notation, express each of the five probabilities defined above using proper probability notation.

   - D: person has the disease
   - H: person is healthy
   - $+$: the test is positive
   - $-$: the test is negative

2. Some congenital disorders are caused by errors that occur during cell division, resulting in the presence of additional chromosome copies. Trisomy 21 occurs in approximately 1 out of 800 births. Cell-free fetal DNA (cfDNA) testing is one commonly used way to screen fetuses for trisomy 21. The test sensitivity is 0.98 and the specificity is 0.995. Calculate the PPV and NPV of the test.
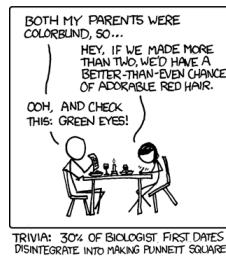
White Cats

The genes that cause cats to be white are linked to genes for eye color and deafness. Suppose that 30% of white cats have one blue eye, while 10% of white cats have two blue eyes. About 73% of white cats with two blue eyes are deaf and 40% of white cats with one blue eye are deaf. Only 19% of white cats with other eye colors are deaf.

1. Estblish some good notation for this problem. [Note: since this is only about white cats, you can think of white cats as the sample space. There is no need to have notation for the event of being a white cat.]

2. Calculate the prevalence of deafness among white cats.

3. Given that a white cat is deaf, what is the probability that it has two blue eyes?

4. Suppose that deaf, white cats have an increased chance of being blind, but that the prevalence of blindness differs according to eye color. Deaf white cats with two blue eyes or two non-blue eyes have probability 0.20 of developing blindness, deaf white cats with one blue eye have probability 0.40 of developing blindness. White cats that are not deaf have probability 0.10 of developing blindness, regardless of their eye color.

   a. What is the prevalence of blindness among deaf, white cats?
   b. What is the prevalence of blindness among white cats?
   c. Given that a cat is white and blind, what is the probability that it has two blue eyes?