Grade 7/8 Math Circles

March 4 & 5 & 6 & 7, 2024

The Bayesian Clinical Diagnostic Model - Problem Set Solutions

1. Let's review some terminology.

   (a) Which of the following test results could you receive if you actually have the disease (*circle some of the terms below*)?

   false positive   false negative   true positive   true negative

   (b) Which of the following test results could you receive if you don't actually have the disease (*circle some of the terms below*)?

   false positive   false negative   true positive   true negative

   (c) If you actually have the disease, the chance that the diagnostic test correctly detects that you have the disease is called (*circle one of the terms below*):

   specificity   sensitivity

   (d) If you don't actually have the disease, the chance that the diagnostic test correctly detects that you don't have the disease is called (*circle one of the terms below*):

   specificity   sensitivity

   (e) Another name for the specificity of a test is: ____________________

   (f) Another name for the sensitivity of a test is: ____________________

Solution:

(a) If you actually have the disease, you could receive a true positive or a false negative.

(b) If you don't actually have the disease, you could receive a true negative or a false positive.

(c) If you actually have the disease, the chance that the diagnostic test correctly detects that you have the disease is called the sensitivity of the test.

(d) If you don't actually have the disease, the chance that the diagnostic test correctly detects that you don't have the disease is called the specificity of the test.

(e) Another name for the specificity of the test is the true negative rate.

(f) Another name for the sensitivity of the test is the true positive rate.

2. Check your understanding.

(a) In a fixed population of people who either have a disease or don't, increasing the number of true negatives will definitely decrease which of the below?

false positive   false negative   true positive   true negative

(b) In a population of people who either have a disease or don't, a test that yields a high number of true positives relative to false negatives has a high ________ (*circle one of the terms below*).

specificity   sensitivity

(c) Which of the two properties of a test can be important in reducing unnecessary psychological impacts of social stigma (*circle the best answer below*)?

specificity   sensitivity

(d) Which quantity or quantities are influenced by the prevalence of disease in a given population?

specificity   sensitivity   positive predictive value

(e) Suppose that the prevalence of a disease is low (about 1%). Suppose that the specificity of a test is 75% and that the sensitivity of the test is 90%. Which of the following adjustments to the positive predictive value would result in the largest increase to the positive predictive value?

i. Raising the sensitivity of the test to 100% and keeping the specificity of the test at 90%.

ii. Keeping the sensitivity of the test at 75% and raising the specificity of the test to 100%.

Solution:

(a) In a fixed population of people who either have a disease or don't, increasing the number of true negatives will definitely decrease the number of false positives. This is because there are a fixed number of people with the disease in the population. If we can increase the specificity of the test (and thus the number of true negatives), then the number of false positives will decrease.

(b) A test that yields a high number of true positives relative to false negatives has a high sensitivity.

(c) The positive predictive value is influenced by the prevalence of disease, whereas the specificity and sensitivity are inherent properties of the test that are not influenced by the prevalence of disease.

(d) Recall that the positive predictive value is the proportion of true positives out of all positives, $\frac{TP}{TP+FP}$. By increasing the test's sensitivity from 90% to 100%, we can increase the number of true positives, which would increase the positive predictive value overall. However, by increasing the test's specificity from 75% to 100%, we would eliminate false positives entirely, which would make the positive predictive value equal to 100%. The positive predictive value would never be 100% with an increase to the test's sensitivity alone, because there would still be some false positives in the denominator.

3. A diagnostic test results in

- 5 true positives
- 2 false positives
- 8 true negatives
- 3 false negatives

Based solely on these results, calculate the most probable

(a) prevalence of the disease
(b) sensitivity of the test
(c) specificity of the test
(d) positive predictive value of the test

Solution:

(a) To determine the prevalence of the disease, we need to find out how many people have the disease and how many people there are in total. The number of people with the disease is the sum of the number of true positives and the number of false negatives, which is $5 + 3 = 8$. The total number of people in the population is

$5+2+8+3=18$. Therefore, the prevalence is

$$\text{Prevalence} = \frac{TP+FN}{TP+FP+TN+FN} = \frac{8}{18} = \frac{4}{9} = 44.\overline{4}\%$$

(b) The sensitivity of the test is

$$\text{sensitivity} = \frac{TN}{FP+TN} = \frac{8}{2+8} = \frac{8}{10} = \frac{2}{5} = 80\%$$

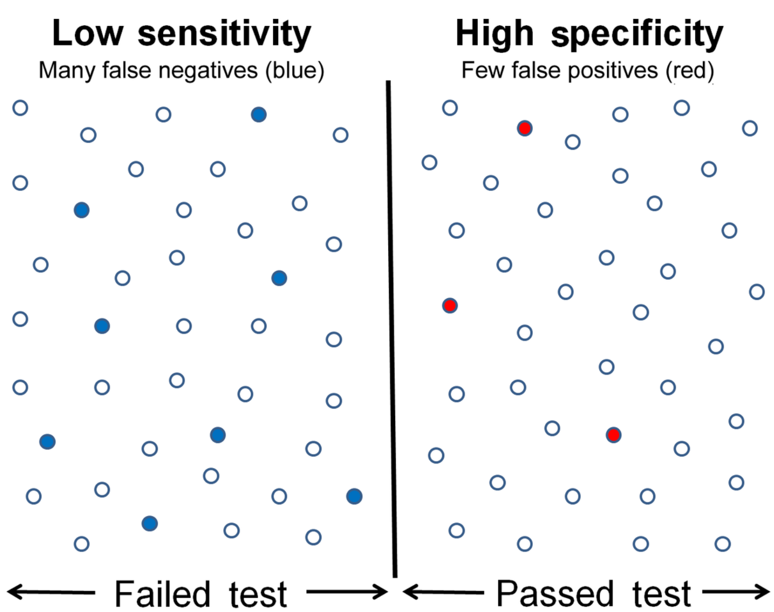
(c) The specificity of the test is

$$\text{specificity} = \frac{TP}{TP+FN} = \frac{5}{5+3} = \frac{5}{8} = 62.5\%.$$

(d) The positive predictive value of the test is

$$\text{Positive predictive value} = \frac{TP}{TP+FP} = \frac{5}{5+2} = \frac{5}{7} = 71.\overline{428571}\%$$

4. Interpret the following diagram[1] of people who have passed or failed a diagnostic test. Calculate the sensitivity and specificity of the test based on the results shown in the diagram.



[1] **Diagram source**: Rmostell, CC0, via Wikimedia Commons

Solution: In this case we have

- There are 8 false negatives (blue dots on the left).
- There are 32 true negatives (uncoloured dots on the left).
- There are 3 false positives (red dots on the right).
- There are 37 true positives (uncoloured dots on the right).

Therefore, the sensitivity is

$$TPR = \frac{TP}{TP+FN} = \frac{37}{37+8} = \frac{37}{45} \approx 82.\overline{2}\%.$$

Also, the specificity is

$$TNR = \frac{TN}{TN+FP} = \frac{32}{32+3} = \frac{32}{35} \approx 91.\overline{428571}\%.$$

5. The sensitivity of a test is 50% and the specificity of the test is 80%. In a group of 1000 people who are tested, 50 actually have the disease.

   (a) What is the prevalence of the disease?

   (b) Predict how many of the 1000 people tested are true positives.

   (c) Predict how many of the 1000 people tested are false positives.

   (d) Predict how many of the 1000 people tested have a positive test result (either a true positive or a false positive).

   (e) What is the positive predictive value of the test?

Solution:

(a) The prevalence of the disease is

$$\begin{aligned}\text{Prevalence} &= \frac{\#\text{ people with the disease}}{\#\text{ people in the population}}\\ &= \frac{50}{1000}\\ &= 5\%\end{aligned}$$

(b) The expected number of true positives is given by the formula (sensitivity $\times$

people with the disease). This gives the number of true positives as

$$\begin{aligned} TP &= \text{sensitivity} \times \#\text{ people with the disease} \\ &= 50\% \times 50 \\ &= 25 \end{aligned}$$

(c) The specificity is the proportion of true negatives out of all true negatives and false positives. Therefore, (1 − specificity) is the fraction of people without the disease who are false positives. So the expected number of false positives is given by the formula (1 − specificity) × (# people without the disease). Using this formula, the number of false positives is

$$\begin{aligned} FP &= (1 - \text{specificity}) \times \#\text{ people without the disease} \\ &= (1 - 0.80) \times (\#\text{ people in the population - }\#\text{ people with the disease}) \\ &= 0.20 \times (1000 - 50) \\ &= 0.20 \times 950 \\ &= 190 \end{aligned}$$

(d) The number of people who have a positive test result is $TP + FP = 25 + 190 = 215$.

(e) Now that we know the number of true positives and false positives (from parts (b) and (c) of the question), we can calculate the positive predictive value as follows.

$$\text{Positive predictive value} = \frac{TP}{TP + FP} = \frac{25}{25 + 190} = \frac{25}{215} \approx 11.63\%.$$

6. Suppose that a test for a certain condition has a sensitivity of 100% and a specificity of 95%. Out of 1000 people tested, 145 test positive and the rest test negative. How many true positives, false positives, true negatives and false negatives are there expected to be?

Solution: Since the sensitivity is 100%, then there are no false negatives (everyone who has the disease gets a true positive test result). Since the specificity is 95%, then the probability that someone without the disease gets a true negative is 95%. There are 145 positives so there must be $1000 - 145 = 855$ negatives. Since there are no false negatives, then all 855 of these negatives are true negatives. This means that 855 is 95% of the people without the disease. So the total number of people without the disease is $\frac{855}{0.95} = 900$, and so there are $900 - 855 = 45$ false positives. The question states that there are a total of 145 positives, and so there must be $145 - 45 = 100$ true positives.

The expected number of true positives is 100; false positives is 45; true negatives is 855; false negatives is 0.

7. Test A is performed on exactly 1 individual who actually has the disease, returning a true positive test result and exactly 2 individuals who don't actually have the disease, both returning a true negative test result. Test B is performed on 360 individuals who actually have the disease and 370 individuals who don't actually have the disease. When the test results come back, 349 of the 360 individuals who have the disease receive a positive test result, with the rest receiving a negative test result, and 350 out of the 370 individuals who don't actually have the disease receive a negative test result, with the rest receiving a positive test result.

   (a) Calculate the sensitivity and specificity for each test.

   (b) Based solely on this data, which test is less risky to use? Explain your reasoning.

Solution:

(a) Test A

The sensitivity is

$$\frac{TP}{\text{everyone who actually has the disease}} = \frac{TP}{TP + FN} = \frac{1}{1} = 100\%.$$

The specificity is

$$\frac{TN}{\text{everyone who actually has the disease}} = \frac{TP}{TN + FP} = \frac{2}{2} = 100\%$$

So both the sensitivity and specificity of Test A is 100%, based on these results.

Test B

The sensitivity is

$$\frac{TP}{\text{everyone who actually has the disease}} = \frac{TP}{TP + FN} = \frac{349}{360} = 96.9\bar{4}\%$$

The specificity is

$$\frac{TN}{\text{everyone who doesn't have the disease}} = \frac{TN}{TN + FP} = \frac{350}{370} = 94.\overline{594}\%.$$

So the sensitivity of Test A is $96.9\bar{4}\%$ and the specificity of Test B is $94.\overline{594}\%$, based on these results.

(b) Even though Test A has both a higher sensitivity and specificity than Test B, Test B is less risky to use based on this data. We are pretty confident that the actual sensitivity and specificity for Test B is around the values we calculated, since there are a large number of data points. On the other hand, we are less confident that the sensitivity and the specificity for Test A will be close to 100%: the sensitivity and specificity of test A might be far away from these values!

8. The *false positive rate* (FPR) or *fallout* of a test is the probability that someone without the disease is not correctly identified by the test (i.e., they receive a false negative).

   (a) Suppose that a test has a sensitivity of 30% and a specificity of 96%. What is the false positive rate of the test?

   (b) Suppose that a test results in 15 false positives and 60 true negatives. What is the false positive rate of the test?

   (c) For a test of a certain ailment, 5 out of every 50 people who don't have the ailment test positive. What is the false positive rate of the test?

   (d) The prevalence of a disease is 1 in 1000. The false positive rate of a test for the disease is 5%, and the test never fails to detect someone who really has the disease. If someone's test result comes back positive, what are the chances that they actually have the disease?

Solution:

(a) The specificity is the probability that someone without the disease is correctly identified by the test. Someone without the disease can either be correctly identified by the test (i.e., they are a true negative) or incorrectly identified by the test (i.e., they are a false positive) when they take the test. Therefore, the fallout of a test is $(1 - \text{specificity})$. In this case, the fallout of the test is $100\% - 96\% = 4\%$.

(b) The false positive rate (or fallout) of the test is

$$FPR = \frac{FP}{\text{everyone who doesn't have the disease}} = \frac{FP}{TN + FP} = \frac{15}{60 + 15} = \frac{15}{75} = 20\%.$$

(c) The false positive rate (or fallout) of the test is

$$\begin{aligned} FPR &= \frac{FP}{\text{everyone who doesn't have the disease}} \\ &= \frac{5}{50} \\ &= 10\%. \end{aligned}$$

(d) Recall that the probability that someone who receives a positive test result actually has the disease is called the positive predictive value, so this is what we are trying to find.

The prevalence of the disease is 1 in 1000 (which is 0.1% when expressed as a percentage). Also, everyone who actually has the disease is detected by the test, so the sensitivity is 100%.

In a group of 1000 people, we'd expect 1 true positive. (The choice to have specifically 1000 people in our population is arbitrary.) As well, the number of false positives in our group of 1000 people is the number of people who don't have the disease multiplied by the fallout. This number would be $999 \times 5\% = 49.95$. Even though this is a fraction, we can use it without rounding in the formula for the positive predictive value to calculate the exact expected positive predictive value.

Putting everything in the formula, we get

$$\begin{aligned}
\text{Positive predictive value} &= \frac{TP}{TP + FP} \\
&= \frac{1}{1 + 49.95} \\
&= \frac{1}{50.95} \\
&\approx 1.96\%
\end{aligned}$$

Despite the low fallout (false negative rate) and therefore high specificity, the chance that someone with a positive test result actually has the disease is quite low due to the low prevalence!