BMI 881 - Homework 2

Tim Gruenloh

November 2024

1 Question 1

If a test has sensitivity = 80% and specificity 80% and the prevalence of the disease is 9/100,000, what is the positive predictive value (aka "precision") of the test?

$$\frac{0.8\left(\frac{9}{100000}\right)}{0.8\left(\frac{9}{100000}\right)+0.2\left(\frac{9}{100000}\right)} = 0.00036$$

The precision would be 0.036%

2 Question 2

Suppose sensitivity = specificity. What would they have to be to achieve positive predictive value = 50% when prevalence is 9/100,000?

Let x be our sensitivity and specificity, and $p = \frac{9}{100{,}000}$ be our prevalence. Then our positive predictive value (PPV) is

$$\begin{aligned} PPV &= \frac{xp}{xp+(1-x)(1-p)} \\ 0.5 &= \frac{xp}{xp+1-p-x+xp} \\ 0.5(2xp+1-x-p) &= xp \\ 0.5x+0.5p &= 0.5 \\ x &= 1-p \end{aligned}$$

So we would need to have sensitivity and specificity at $1-\frac{9}{100{,}000} = 0.99991$

3 Question 3

Comment on these results in relation to the precision values provided in Table 2 of Wang et al. (2019).

In the paper, they have sensitivity and specificity at slightly over 80%, which means that, given the population level prevalence, their precision should be very small. However, they report a precision of 0.571, which doesn't make sense given the calculations in questions 1 and 2.

This can be attributed to the fact that the prevalence in their test set (and also train set) (about 20%) is much higher than the actual prevalence of the disease. This paper and assignment are a good reminder to make sure that the data you train your model on is representative of the population you wish to treat.