Calculating PPV in terms of sensitivity, specificity and prevalence

Calculating the PPV in terms of sensitivity, specificity, and prevalence

In some studies, you may have to compute the Positive predictive value (PPV) from the sensitivity, specificity and prevalence. Note that reviewing this reading will help you answer one of the quizzes at the end of this week!

Rewriting PPV

$PPV = P(pos|\hat{pos})$.

(pos is "actually positive" and $\hat{pos}$ is "predicted positive").

By Bayes rule, this is

$PPV = \frac{P(\hat{pos}|pos) \times P(pos)}{P(\hat{pos})}$

For the numerator:

$Sensitivity = P(\hat{pos}|pos)$. Recall that sensitivity is how well the model predicts actual positive cases as positive.

$Prevalence = P(pos)$. Recall that prevalence is how many actual positives there are in the population.

For the denominator:

$P(\hat{pos}) = TruePos + FalsePos$. In other words, the model's positive predictions are the sum of when it correctly predicts positive and incorrectly predicts positive.

The true positives can be written in terms of sensitivity and prevalence.

$TruePos = P(\hat{pos}|pos) \times P(pos)$, and you can use substitution to get

$TruePos = Sensitivity \times Prevalence$

The false positives can also be written in terms of specificity and prevalence:

$FalsePos = P(\hat{pos}|neg) \times P(neg)$

$$1 - specificity = P(pos|neg)$$

$$1 - prevalence = P(neg)$$

PPV rewritten:

If you substitute these into the PPV equation, you'll get

$$PPV = \frac{sensitivity \times prevalence}{sensitivity \times prevalence + (1 - specificity) \times (1 - prevalence)}$$