

Lab 3. Sensitivity, Specificity, PPV, NPV, ROC

In this lab, we will work on the exercise 15 and 16 in chapter 6 (probability) of Pagano and Gauvreau. The purpose is to further illustrate the definition and calculation of sensitivity, specificity, PPV, NPV, and ROC.

15. The following data are taken from a study investigating the use of a technique called radionuclide ventriculography as a diagnostic test for detecting coronary artery disease.

Test	Disease		Total
	Present	Absent	
Positive	302	80	382
Negative	179	372	551
Total	481	452	933

(a) *What is the sensitivity of radionuclide ventriculography in this study? What is its specificity?*

Solution: Recall the definition of sensitivity: the probability of a positive test result given that the individual really has the disease. From the table we can see that there are 481 subjects really having the disease, among which 302 had positive test result. Therefore, the sensitivity of radionuclide ventriculography is:
Sensitivity = $302/481 = 62.8\%$ or 0.628.

Similarly, the specificity is defined as the probability of a negative test result given that the individual is disease-free. There are 452 disease-free subjects, among which 372 had negative test result. Therefore, the specificity is:
Specificity = $372/452 = 82.3\%$ or 0.823.

(b) *For a population in which a prevalence of coronary artery disease is 0.10, calculate the probability that an individual has the disease given that he or she tests positive using radionuclide ventriculography.*

Solution: By definition, what the problem asks for is the positive predictive value (PPV) of the test. Let's define event A to be having the disease and event B to be having a positive test result. By definition of prevalence we have $\Pr(A) = 0.10$. (Note that we are no longer using the above table). What we want is $\text{PPV} = \Pr(A|B)$. Recall the Bayes's theorem:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(B|A) \Pr(A)}{\Pr(B \cap A) + \Pr(B \cap \bar{A})}$$

$$\Pr(A) =$$

$$\Pr(B|A) =$$

$$\Pr(B \cap A) = \Pr(B|A) \Pr(A) =$$

$$\Pr(B \cap \bar{A}) = \Pr(B|\bar{A}) \Pr(\bar{A}) = (1 - \Pr(\bar{B}|\bar{A}))(1 - \Pr(A)) =$$

$$\text{Therefore, PPV} = \Pr(A|B) =$$

Alternative solution: There is a more concrete way to calculate $\text{PPV} = \Pr(A|B)$. Let's assume the total number of subjects in the study is n (or any positive integer). Then we can construct a table similar to the one given above, only with n in each cell.

Test	Disease		Total
	Present	Absent	
Positive			
Negative			
Total			

$$\text{So PPV} = \Pr(A|B) =$$

(c) What is the predictive value of a negative test?

Solution: By definition, negative predictive value (NPV) is $\Pr(\bar{A}|\bar{B})$. By Bayes's theorem:

$$\Pr(\bar{A}|\bar{B}) = \frac{\Pr(\bar{A} \cap \bar{B})}{\Pr(\bar{B})} = \frac{\Pr(\bar{B}|\bar{A}) \Pr(\bar{A})}{\Pr(\bar{B} \cap A) + \Pr(\bar{B} \cap \bar{A})}$$

$$\Pr(\bar{A}) =$$

$$\Pr(\bar{B}|\bar{A}) =$$

$$\Pr(\bar{B} \cap A) = \Pr(\bar{B}|A) \Pr(A) = (1 - \Pr(B|A)) * \Pr(A) =$$

$$\Pr(\bar{B} \cap \bar{A}) = \Pr(\bar{B}|\bar{A}) \Pr(\bar{A}) =$$

Therefore,

$$\text{NPV} = \Pr(\bar{A}|\bar{B}) =$$

The alternative solution is similar to (b) and is left as exercise.

16. The table below displays data taken from a study comparing self-reported smoking status with measured serum cotinine level. As part of the study, cotinine level was used as a diagnostic tool for predicting smoking status; the self-reported status was considered to be true. For a number of different cutoff points, the observed sensitivities and specificities are given below.

Cotinine level (ng/ml)	Sensitivity	Specificity
5	0.971	0.898
7	0.964	0.931
9	0.960	0.946
11	0.954	0.951
13	0.950	0.954
14	0.949	0.956
15	0.945	0.960
17	0.939	0.963
19	0.932	0.965

(a) As the cutoff point is raised, how does the probability of a false positive result change? How does the probability of a false negative result change?

Solution: False positive rate = $1 - \text{specificity}$. From the table we can see that, as the cutoff point is raised, the specificity increases, so the false positive rate decreases. False negative rate = $1 - \text{sensitivity}$, and it increases as the cutoff point is raised. Note that false positive rate and false negative rate always change in the opposite direction for any given test.

(b) Use these data to construct a receiver operator characteristic curve.

Solution: We will use Stata to generate the ROC curve. By definition, ROC curve is a plot of sensitivity against false positive rate for a series of cutoff points. So we first need to generate false positive rate in Stata using the command:

generate fpr = 1 - specificity

Then use the following command to plot sensitivity against false positive rate and connect the dots by step function.

```
twoway (line sensitivity fpr, connect(stairstep)), ytitle(Sensitivity)
xtitle(FPR (1-specificity)) title(ROC curve on Cotinine level-based
test)
```

Theoretically sensitivity and false positive rate can range between 0 and 1. The data in this study only span a small range. To get a bigger picture of the ROC curve, we rescale the y and x axis to [0, 1]. The following command generates the rescaled ROC curve with a 45 degree reference line:

```
twoway (line sensitivity fpr, connect(stairstep)), ytitle(Sensitivity)
ylabel(0(0.1)1) xtitle(False positive rate (1 - specificity)) xlabel(0(0.1)1)
title(Rescaled ROC curve on Cotinine level-based test) legend(off) ||
function y=x, range(0 1)
```

(c) Based on this graph, what value of serum cotinine level would you choose as an optimal cutoff point for predicting smoking status? Why?

Solution: Usually the optimal cutoff point on a ROC curve is chosen as the one that is closest to the upper-left corner. This cutoff point gives the best combination of sensitivity and specificity. In reality, the cost of false positive result and false negative result has to be taken into consideration when choosing the optimal cutoff point for a given diagnostic test. In this particular study, based on the ROC curves you just generated, what do you think about the choice of the optimal cutoff point?