Data Science

Working with Data

This questioner article has addressed the following subjects.

* Probability
  + true positive rate
  + true negative rate
* Bayes' theorem
* Python libraries: NumPy, SciPy

**Task I: Probability**

Consider a medical diagnostic scenario where a patient undergoes a test for a rare disease. The disease has a prevalence rate of 1% in the population. The diagnostic test has a sensitivity (true positive rate) of 95% and a specificity (true negative rate) of 90%.

**Part I:**

Calculate the following probabilities using Python:

The probability that a person who tests positive for the disease actually has the disease (positive predictive value).

The probability that a person who tests negative for the disease does not have the disease (negative predictive value).

Include the Python code you used for the calculations.

**Python code:**

A computer code with many colorful text

Description automatically generated with medium confidence

**Part II:**

The patient undergoes the diagnostic test and receives a positive result. Using Bayes' theorem and Python, compute the probability that the patient actually has the disease. Show your Python code and step-by-step calculations, and explain your reasoning.

**Bayes’ theorem:**

A black text on a white background

Description automatically generated

**Python code:**

A screenshot of a computer code

Description automatically generated

**Manually calculating the Probability that the person actually has the disease.**

A close-up of a musical note

Description automatically generated

A close-up of a white background

Description automatically generated

**A close-up of a white board

Description automatically generated**

**A close-up of a white background

Description automatically generated**

**The probability that the patient actually has the disease:**

Using Python: 0.008766

Manual calculation: 0.0963

**Part III:**

Discuss the implications of your results in part (b) for medical decision-making and public health policies. Consider the importance of understanding Bayes' theorem and probability concepts when interpreting diagnostic test results. Propose strategies to improve the accuracy of the diagnostic test, and explain the potential impact of these strategies on the probabilities calculated in part (a). Discuss how Python libraries, such as NumPy or SciPy, can be used to aid in the analysis and improvement of diagnostic tests in the context of data science.

Typically, diagnostic tests are described by their sensitivity and specificity.

Sensitivity is the capacity of a test to identify a diseased person as positive. A highly sensitive test produces fewer false negative results and, as a result, misses fewer illness cases. As per our question, the sensitivity rate is 95%. It means, the identifying a person has illness is very high. Since this is a higher value, it means that the test has missed a few cases only.

A test's specificity is measured by its capacity to label as negative someone who doesn't have a condition. Few false positive findings come from a highly precise test. In the question, the specificity rate is 90% and it means the test have been properly conducted with most accurate results.

Understanding these ideas makes it easier to evaluate a test's correctness and high likelihood of producing both false positives and false negatives results.

The Bayes theorem is based on the idea that the usefulness of diagnostic tests is influenced by both the prevalence of the disease in the test population as well as sensitivity and specificity., the properties of the diagnostic test. On the other hand, based on the likelihood that a person will actually have the illness, we can modify the test findings using Bayes' theorem. Using this, we can calculate positive predictive value (PPV) and negative predictive value (NPV) of this diagnostic test. PPV indicates the likelihood that a test result is positive, whereas NPV indicates the likelihood that a test result is negative and these are more informative when it comes to the values of sensitivity and specificity, because these 2 consider the prevalence of the disease in the population.

Therefore, The Bayes theorem can be used to forecast the outcome of various tests and how they will affect patient care. Therefore, in order to do these predictions more accurately, we can use Python libraries, such as NumPy or SciPy. Moreover, by involving more patients on diagnostic tests, the accuracy of test results can be increased significantly.

|  |
| --- |
| (New York State Department of Health, n.d.) References New York State Department of Health, n.d. *New York State Department of Health.* [Online]  Available at: https://www.health.ny.gov/diseases/chronic/discreen.htm [Accessed 10 September 2023].  (National Library of Medicine, 2021) References National Library of Medicine, 2021. *National Library of Medicine National Center for Biotechnology Information.* [Online]  Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8156826/ [Accessed 10 September 2023].  (ScienceDirect, 2020) References ScienceDirect, 2020. *ScienceDirect.* [Online]  Available at: https://www.sciencedirect.com/science/article/pii/S0895435620312257 [Accessed 10 September 2023]. |