Shweta_Shinde_Math_Assignment10

November 30, 2024

1 Shweta Ajay Shinde 017548687

MSDA, SJSU , Data 220- Math Method for DA

Q1.Suppose there is a rare disease that affects 1 in 10,000 people in a population. There is a diagnostic test for this disease that has the following characteristics:

- The probability of a true positive (the test correctly identifies a disease when it is present) is 99%.

- The probability of a true negative (the test correctly identifies no disease when it is absent) is 95%.

- False positives and false negatives can occur.

Introduction to Bayes' Theorem For this we will use Bayes' Theorem Bayes' Theorem for this scenario is:

```
- P(Desiese | Positive Test) =  P(Desiese | Positive Test) * P(Desiese)
                                --------------------------------------------
                                              P(Positive Test)
```

Where:

- (Disease) = P(Disease) is the prior probability of having the disease (1 in 10,000).

- P(Positive Test Disease) is the sensitivity, i.e., the probability of a true positive (99%).

- P(Positive Test) is the total probability of testing positive, which can be broken down into two parts:

  ```
  - The true positives:  P(Positive Test Disease)×P(Disease)
  ```

  ```
  - The false positives: P(Positive Test No Disease)×P(No Disease)
  ```

  Where:

- P(Positive Test No Disease) is the probability of a false positive, which is 1−True Negative Rate, i.e., 5% (because the true negative rate is 95%).

- P(No Disease) is the probability of not having the disease, which is 1−P(Disease).

The Formula for

- P(Positive Test) = P(Positive Test Disease) × P(Disease) + P(Positive Test No Disease) × P(No Disease)

Python Code

```python
[4]: def posterior_probability(test_positive, disease_prevalence,
     ↪true_positive_rate, true_negative_rate):

         #p_disease = disease_prevalence
         #p_positive_given_disease = true_positive_rate

         # Probability of not having the disease
         p_no_disease = 1 - disease_prevalence

         # False positive rate (1 - specificity)
         p_positive_given_no_disease = 1 - true_negative_rate

         # Probability of a positive test result (Total Probability)
         p_positive_test = (true_positive_rate * disease_prevalence) +
     ↪(p_positive_given_no_disease * p_no_disease)

         # Applying Bayes' Theorem to calculate posterior probability
         posterior = (true_positive_rate * disease_prevalence) / p_positive_test

         return posterior
```

1.0.1 Explanation:

1.0.2 Bayes' Theorem and Posterior Probability

Bayes' Theorem helps us calculate the probability of having a disease given a positive test result, which is called the posterior probability. It uses prior knowledge (like the disease prevalence) and the test's performance (sensitivity and specificity).

- Prior Probability: This is the probability of having the disease before the test, based on how common the disease is (1 in 10,000 people in this case).

- Sensitivity: The probability that the test correctly identifies a person with the disease (99% in this case).

- Specificity: The probability that the test correctly identifies a person without the disease (95% here).

Despite the high sensitivity and specificity, the posterior probability is still low due to the rarity of the disease. Even if the test is accurate, there will be some false positives (people without the disease who test positive), and because the disease is so rare, these false positives have a significant impact on the results.

Function Definition: The function posterior_probability is defined with four inputs:

- test_positive: True positive rate (sensitivity).

- disease_prevalence: Prior probability of having the disease.
- true_positive_rate: The sensitivity of the test.
- true_negative_rate: The specificity of the test (probability of correctly identifying no disease).

1.0.3 An example of how to use your function with the provided scenario data (3 pts).

Scenario Data:

- The disease prevalence is 1 in 10,000 people (i.e., 0.0001 probability of having the disease).
- The true positive rate (sensitivity) is 99% (i.e., 0.99 probability of a correct positive result if the disease is present).
- The true negative rate (specificity) is 95% (i.e., 0.95 probability of a correct negative result if the disease is not present).

We will pass all the know values (as given above) as parameter to the defined function **posterior_probability**

```
[5]: # Scenario Data
     disease_prevalence = 1 / 10000  # Disease prevalence: 1 in 10,000
     true_positive_rate = 0.99       # Sensitivity: 99%
     true_negative_rate = 0.95       # Specificity: 95%

     # Call the function to calculate the posterior probability of having the␣
      ↪disease given a positive test result
     posterior_prob = posterior_probability(true_positive_rate, disease_prevalence,␣
      ↪true_positive_rate, true_negative_rate)

     # Print the result
     print(f"The posterior probability of having the disease given a positive test␣
      ↪result is: {posterior_prob:.6f}")
```

The posterior probability of having the disease given a positive test result is: 0.001976

The function computes the posterior probability using these values and applies Bayes' Theorem. The result is approximately **1.98%** —this means that even after testing positive, there's still only about a **1.98%** chance that the person actually has the disease due to its rarity and the possibility of false positives.

[ ]: