HIV - Lab

|  |  |  |  |
| --- | --- | --- | --- |
| Predicted | Actual | | |
|  | Negative | Positive |
| Negative | True Negative | False Negative |
| Positive | False Positive | True Positive |

**Sensitivity**: probability of a true positive. Ex: the probability of testing positive when a patient has breast cancer**.**

**Specificity**: probability of a true negative. Ex: the probability of testing negative when a patient doesn't have breast cancer

**Conditional probabilities and HIV tests**

In order to calculate the how well HIV tests work, the following samples were taken from known HIV+ patients and HIV- patients:

|  |  |  |
| --- | --- | --- |
| Test result | Known HIV+ | Known HIV free |
| Positive | 990 | 15 |
| Negative | 10 | 985 |

P(T+ ) = P(T+ | HIV + ) P( HIV+ ) + P(T+|HIV-) P(HIV-)

a) What is the sensitivity of the HIV test?

P(T + | HIV+) = P(T and HIV) /P(HIV+) = (990/2000) / (1000/2000)

b) What is the specificity of the HIV test?

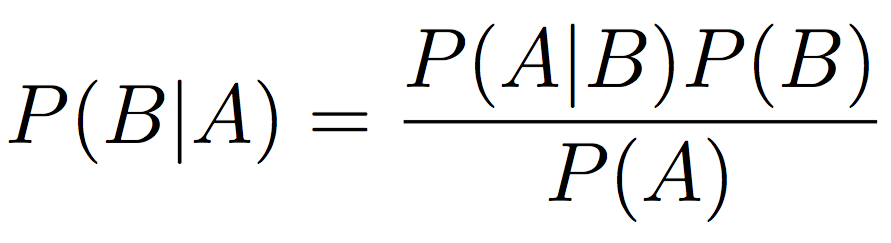
P(T- | HIV-) = P(T- and HIV-)/P(HIV-) = (985/2000) / (1000/2000)

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

c) A patient goes into a clinic and gets an HIV test. She is dismayed to learn that it comes back positive. Now we are going to calculate the probability that she actually is carrying the virus. Put another way, we want to know:

P(HIV+ | T+)

In order to calculate this, we will use Bayes rule. As a reminder:



Express P(HIV+ | T+) in terms of P(T+), P(HIV+) and P(T+| HIV+)

P(HIV+ |T+) = P(T+| HIV+ ) P(HIV+) /

[P(T+| HIV+) P(HIV+) + P(T+|HIV-) P(HIV-)]

d) As you can see, we need to know P(HIV+), or the prior probability that our patient had HIV before we knew anything about her test. A good prior to use in this case is the prevalence of the disease, or the fraction of people that are infected with the virus.

Luckily we have data of HIV prevalence in different regions of the world:

|  |  |
| --- | --- |
| **Region** | **Adult Prevalence** |
| **Sub-Saharan Africa** | **0.059** |
| **South and South East Asia** | **0.006** |
| **East Asia** | **0.001** |
| **Latin America** | **0.005** |
| **North America** | **0.008** |
| **Western and Central Europe** | **0.003** |
| **Eastern Europe and Central Asia** | **0.009** |
| **Middle East and North Africa** | **0.002** |
| **Caribbean** | **0.012** |
| **Oceania** | **0.004** |

Load the data from the file: region\_prevalence.csv

e) Add a column to your data frame where you calculate P(T+) or the total probability of observing a positive test in each region of the world.

P(T+) = P(T+ | HIV+) P(HIV+) + P(T+ | HIV-) P(HIV-)

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

f) Now you have all the tools to calculate the posterior probabilities of patients in different regions being HIV+ given that they had a positive HIV test. (Recall, posterior: P(HIV+ | T+))

Add a column to your data frame that shows these posterior probabilities

Our patient is from Lebanon - should she be very worried that she is HIV+?

What is her posterior probability of carrying the virus?

P(HIV+ | T + ) = 0.1168142

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

g) Is the above result surprising?

YES quite surprising. If you say it out loud: you only have a 11% chance of actually contracting HIV even if you test positive

**Challenge:**

h) In order to be sure of the diagnosis, she goes to a different clinic and gets a second HIV test. Write the probability, “P(?)”, equal to the posterior probabilities after getting a second independent, positive HIV test. Where T1 and T2 are the two independent HIV tests.

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

i) Compute another column in your data frame showing the posterior probabilities for each region after getting a second independent, positive HIV test. If our Lebanese patient got a second positive test, what is her posterior probability of carrying the virus? Should she be worried now?

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

P(HIV + | T1 and T2) = P(T1 and T2 | HIV+) P(HIV+) / P(T1 and T2)

j) What about a patient from Sub-Saharan Africa?

NOTE: As you apply Bayes' rule you might notice a pattern like the one above. As you collect more data, your outcomes become less sensitive to your priors or your original assumptions. This is one reason why people like big data.

Conditional Probability and HIV Testing: A Real-World Example

Chad R. Bhatti and Jennifer L. Wightman

*The American Statistician*

Vol. 62, No. 3 (Aug., 2008), pp. 238-241