Bayes’ Theorem
& Diagnostic Tests
Screening Tests

Some Questions

- If you test positive for HIV, what is the probability that you have HIV?
- If you have a positive mammogram, what is the probability that you have breast cancer?
- Given that you have breast cancer, what is the probability that you will live?

Example: Coronary artery disease

<table>
<thead>
<tr>
<th></th>
<th>Present $D^+$</th>
<th>Absent $D^-$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive $T^+$</td>
<td>815</td>
<td>115</td>
<td>930</td>
</tr>
<tr>
<td>Negative $T^-$</td>
<td>208</td>
<td>327</td>
<td>535</td>
</tr>
<tr>
<td>Total</td>
<td>1023</td>
<td>442</td>
<td>1465</td>
</tr>
</tbody>
</table>

[From the book, Medical Statistics page 30, and the 2x2 table from data of Weiner et al (1979)]

The disease Prevalence in these patients
$P(D^+) = \frac{1023}{1465} = .70$

Predictive Value of a Positive Test: The probability of having the disease given that a person has a positive test is given by:

$$P(D^+|T^+) = \frac{P(D^+ \cap T^+)}{P(T^+)} = \frac{815/1465}{930/1465} = \frac{815}{930} \approx .88$$

Predictive Value of a Negative Test:

$$P(D^-|\neg T^-) = \frac{P(D^- \cap \neg T^-)}{P(\neg T^-)}$$
Screening Tests

<table>
<thead>
<tr>
<th>Present $D^-$</th>
<th>Absent $D^+$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive $T^+$</td>
<td>815</td>
<td>115</td>
</tr>
<tr>
<td>Negative $T^-$</td>
<td>208</td>
<td>327</td>
</tr>
<tr>
<td>Total</td>
<td>1023</td>
<td>442</td>
</tr>
</tbody>
</table>

Things to notice: (Prevalence)

\[
P(D^+) = P(T^+ \cap D^+) + P(T^- \cap D^-) = 815/1465 + 208/1465 = 1023/1465 \approx .7
\]

Likewise, (Overall percentage that tested positive)

\[
P(T^+) = P(T^+ \cap D^+) + P(T^+ \cap D^-) = 815/1465 + 115/1465 = 930/1465 \approx .63
\]

Calculate the following:

\[
P(D^-) = \quad\quad
\]

\[
P(T^-) = \quad\quad
\]

\[
P(D^- \cap T^+), P(D^-)P(T^+) = \quad\quad
\]

Sensitivity and Specificity

**Sensitivity:** The probability of a person testing positive, given that (s)he has the disease.

\[
P(T^+ \mid D^+) = \frac{P(T^+ \cap D^+)}{P(D^+)} = \frac{815/1465}{1023/1465} = \frac{815}{1023} \approx .80
\]

**Specificity:** The probability of a person testing negative given that (s)he does not have the disease.

\[
P(T^- \mid D^-) = \frac{P(T^- \cap D^-)}{P(D^-)} = \frac{327/1465}{442/1465} = \frac{327}{442} \approx .74
\]

Note that **sensitivity is not affected by disease prevalence.** For example, if number of people with true coronary artery disease has tripled from 1023 to 3069 so that the new prevalence is now 3069/(1465+3069) = .68, then we should expect that three times as many patients would test positive. Thus 3x815 = 2445 should have a positive result.

The sensitivity would be \(2445/3069 \approx .80\)
Screening Tests

<table>
<thead>
<tr>
<th></th>
<th>Present D+</th>
<th>Absent D-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive T+</td>
<td>815</td>
<td>115</td>
<td>930</td>
</tr>
<tr>
<td>Negative T-</td>
<td>208</td>
<td>327</td>
<td>535</td>
</tr>
<tr>
<td>Total</td>
<td>1023</td>
<td>442</td>
<td>1465</td>
</tr>
</tbody>
</table>

False Negative rate = 1 – sensitivity

\[ P(T^- | D^+) = \frac{P(T^- \cap D^+)}{P(D^+)} = \frac{208/1465}{1023/1465} = \frac{208}{1023} \approx 0.2 \]

False Positive rate = 1 – specificity

\[ P(T^+ | D^-) = \frac{P(T^+ \cap D^-)}{P(D^-)} = \frac{115/1465}{442/1465} = \frac{115}{442} \approx 0.26 \]

**Sensitivity & False Negatives**

For our problem, since the sensitivity is .8, the false negative rate is 1 - .8 = .2.

Interpretation: 20% of the time a person will actually have the disease when the test says that he/she does not.

Likewise, since specificity is .74, the false positive rate is 0.26.

**Sensitivity vs Specificity**

- In a perfect world, we want both to be high.
- The two components have a see-saw relationship.
- Comparisons of tests’ accuracy are done with Receiver Operator Characteristic (ROC) curves.

**ROC**

<table>
<thead>
<tr>
<th>Test Positive Criteria</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.93</td>
<td>0.55</td>
</tr>
<tr>
<td>3</td>
<td>0.81</td>
<td>0.59</td>
</tr>
<tr>
<td>4</td>
<td>0.64</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**About ROC Curve**

- Area under the ROC represents the probability of correctly distinguishing a normal from an abnormal subject on the relative ordering of the test ratings.
- Two screening tests for the same disease, the test with the higher area under its ROC curve is considered the better test, unless there is particular level of sensitivity or specificity is especially important in comparing the two tests.

**How can we get the predictive value of a positive test from sensitivity?**

After all, at least as a patient, we want to know that probability that we have a disease given that we have just tested positive!

**Bayes’ Theorem**
Screening Tests

**Bayes’ Theorem**

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(A \cap B)}{P(A)} \]

Solving the first equation as follows,

\[ P(A \cap B) = P(A|B)P(B) \]

Substituting this in for the second equation, we have

**Bayes’ Theorem**

\[ P(B|A) = \frac{P(A|B)P(B)}{P(A)} \]

Applying this to our formula, we have

\[ P(D^+|T^-) = \frac{P(T^-|D^-)P(D^-)}{P(T^-)} = \frac{(0.8)(0.7)}{0.63} = 0.99 \]

In words, the predictive value of a positive test is equal to the sensitivity (=.8) times prevalence (.7) divided by percentage who test positive (.63).

**Example**

Suppose that 84% of hypertensives and 23% of normotensives are classified as hypertensive by an automated blood-pressure machine. If it is estimated that the prevalence of hypertensives is 20%, what is the predictive value of a positive test? What is the predictive value of a negative test?

Sensitivity = .84 Specificity = 1-.23 = .77

Pretty Good Test!??

\[ PV^+ = \frac{(0.84)(0.2)}{(0.84)(0.2) + (0.23)(0.8)} = \frac{0.168}{0.26} = 0.64 \]

\[ PV^- = 0.95 \]

What if only 2% prevalence?

**Incorporating New Information to Update our Beliefs**

Use Bayes’ Theorem!

Two examples:

■ Misfiling Assistants
■ AIDS Prevalence

**Finding Disease Prevalence in a Specified Population**

◆ Want to know: what percentage of a certain group has HIV?
◆ Use a screening test on a sample of n individuals from the group.
◆ Know the sensitivity and specificity from prior experience.

**Finding Prevalence**

◆ Suppose n+ people test positive.
◆ Using Bayes’ Theorem we can derive the formula:

\[ P(D^+) = \frac{(n^+ / n) - P(T^+|D^-)}{P(T^+|D^+) - P(T^+|D^-)} \]