BIOM5010:
Statistics #2H

Detection Error Trade-off
Training and Test sets

Test Outcomes

- Example:
  - Reality:
    - Patient Has Cancer
    - Patient Has not Cancer = Patient Normal
  - Test Results:
    - Test says Cancer
    - Test says Not Cancer = Test Normal
Test Outcomes

- Example:
  - Reality:
  - Test Results:

  (Test: Cancer)
  Reject $H_0$
  (Cancer) $H_0$ false
  (Normal) $H_0$ true

  (Test: Normal)
  Accept $H_0$
  $TP$
  $FP$
  $FN$
  $TN$

  Reality

  Test

  - TP = True Positive
  - TN = True Negative
  - FN = False Negative
  - FP = False Positive

  How serious FN, FP depends on application

Example:

- Test for cancer gives 4 levels
  - e.g. Measures level of protein in test
    - 1 = least, more likely to indicate no cancer
    - 4 = most, more likely to indicate cancer
  - 8 patients in Control group
  - 4 patient in Experimental group

- Terms
  - Control group
    - do not have the condition or receive sham/pacebo
  - Experimental group
    - Have condition or receive experimental treatment
Test for cancer gives 4 levels
- Control:
  - 1,1,1,1,2,2,3,3
- Experimental:
  - 2,3,3,4

Control: Distribution

Experimental: Distribution

Fraction of group

Test Result
• Use a threshold to make decision from score

**Sensitivity/Specificity**

- **Control:**
  - Threshold: 1.0
  - True Negative (TN)
  - False Positive (FP)
  - False Negative (FN)
  - True Positive (TP)

- **Experimental:**
  - Threshold: 1.0
  - True Negative (TN)
  - False Positive (FP)
  - False Negative (FN)
  - True Positive (TP)
Sensitivity/Specificity

- Need a threshold to make decision from score
- From Distributions Calculate Specificity and Sensitivity

### Specificity

<table>
<thead>
<tr>
<th>Threshold</th>
<th>TN</th>
<th>FP</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>1.5</td>
<td>4</td>
<td>4</td>
<td>0.5</td>
</tr>
<tr>
<td>2.5</td>
<td>6</td>
<td>2</td>
<td>0.75</td>
</tr>
<tr>
<td>3.5</td>
<td>8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4.5</td>
<td>8</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### Sensitivity

<table>
<thead>
<tr>
<th>Threshold</th>
<th>FN</th>
<th>TP</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>1.5</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td>3.5</td>
<td>3</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>4.5</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Detection Error Trade-off (DET)

- From Distributions Calculate DET

<table>
<thead>
<tr>
<th>Threshold</th>
<th>TN</th>
<th>FP</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>1.5</td>
<td>4</td>
<td>4</td>
<td>0.5</td>
</tr>
<tr>
<td>2.5</td>
<td>6</td>
<td>2</td>
<td>0.75</td>
</tr>
<tr>
<td>3.5</td>
<td>8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4.5</td>
<td>8</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Threshold</th>
<th>FN</th>
<th>TP</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>1.5</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td>3.5</td>
<td>3</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>4.5</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: you can always get sensitivity or specificity = 100%
• DET curve or Receiver Operating Characteristic (ROC)
• Common Way to show instrument characteristics
• You can always get any value to 100% by setting the other one to zero.
• To properly compare two curves:
  – Compare at one choice of sensitivity
  – Compare Areas under curve
Comparing DET/ROCs

- To compare curves, need to choose an operating level
- Choose acceptable sensitivity
  - Find corresponding specificity

\[
\text{Sensitivity} = 0.6 \quad \text{Sensitivity} = 0.9
\]

At \(\text{sens} = 0.9\), \(A > B\)

At \(\text{sens} = 0.6\), \(B > A\)
Questions

• Is a FP more serious than a FN?
• How do we go about setting a threshold for an algorithm with an output score?
• Under what conditions (on the distributions) can we have sensitivity = specificity = 100%?
• Suggest some clinical scenarios where we would want to trade sensitivity for specificity? Vice-versa?
• The distributions have no information on the size of the data samples. How can we still calculate sensitivity values?