Prediction

HL Chapter 5 - part 3

Classification tables

How well does the model classify subjects as diseased or non-diseased?

Validity and predictive values

- Calculate sensitivity (Se), specificity (Sp), and positive and negative predictive values (PPV, NPV) for the model

Cutpoint selection

Predicted Outcome	Observed Outcome		
	Present	Absent	Total
Present	True +	False +	
Absent	False -	True -	
Total			

- For each subject we must decide if $\hat{\pi}$ suggests that the outcome is present or that the outcome is absent
- To do this, we must choose a cutpoint, e.g.
  - $\hat{\pi} \geq 0.5$ ➔ predicted outcome present
  - $\hat{\pi} < 0.5$ ➔ predicted outcome absent

Examples

Predicted Outcome	Observed Outcome		
	Present	Absent	Total
Present	True +	False +	
Absent	False -	True -	
Total			

- A subject with $\hat{\pi}$ =0.7 ($\hat{\pi}$ >0.5) and outcome present counts toward the true positives
- A subject with $\hat{\pi}$ =0.7 ($\hat{\pi}$ >0.5) and outcome absent counts toward the false positives

Examples

Predicted Outcome	Observed Outcome		
	Present	Absent	Total
Present	True +	False +	
Absent	False -	True -	
Total			

- A subject with $\hat{\pi}$ =0.3 ($\hat{\pi}$ <0.5) and outcome present counts toward the false negatives
- A subject with $\hat{\pi}$ =0.3 ($\hat{\pi}$ <0.5) and outcome absent counts toward the true negatives

Problems

- Continuous probabilities ($\hat{\pi}$) are reduced to 0/1
- If 0.5 is chosen as the cutpoint,
  - $\hat{\pi}$=0.49 and $\hat{\pi}$=0.01 are considered the same
  - $\hat{\pi}$=0.49 and $\hat{\pi}$=0.51 are considered different
- 0.5 may not be the most appropriate cutpoint

Example: Glow500 data set

Determine best cutpoint for the classification table

```
proc logistic descending data=glow500;
  model fracture=priorfrac momfrac armassist raterisk2
                 height age priorfrac*age
                 momfrac*armassist
        /outroc=rocdat;
run;
```

Creates data set containing sensitivity and 1-specificity for all values of $\hat{\pi}$ in the data set

```
data rocdat;
  set rocdat;
  spec=1-_1mspec_;
run;
```

Calculates specificity
(SAS provides 1-Specificity)

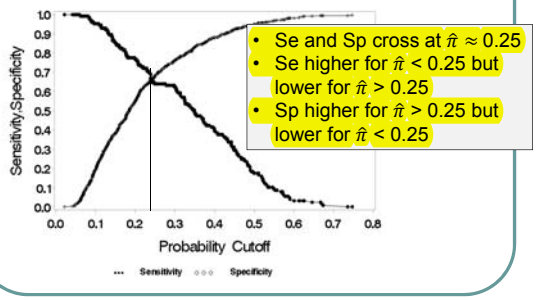
```
axis1 label=(f=swiss h=2.5 'Probability Cutoff') minor=none;
axis2 label=(f=swiss h=2.5 a=90 'Sensitivity,Specificity') minor=none;

goptions FTEXT=swissb HTEXT=2.0 HSIZE=8 in VSIZE=6 in;

symbol1 v=dot i=join c=black h=1;
symbol2 v=diamond i=join c=black h=1;

footnote1 c=black f=special h=1 'J J J' f=swissb h=1.5 '  Sensitivity'
      c=black f=special h=1 '  D D D' f=swissb h=1.5 '  Specificity';
```

```
proc gplot data=rocdat;
  plot (_sensit_ spec)*_prob_
        /overlay haxis=axis1 vaxis=axis2;
run; quit;
```



- Se and Sp cross at $\hat{\pi} \approx 0.25$
- Se higher for $\hat{\pi} < 0.25$ but lower for $\hat{\pi} > 0.25$
- Sp higher for $\hat{\pi} > 0.25$ but lower for $\hat{\pi} < 0.25$

Create the classification table

```
proc logistic descending data=glow500;
  model fracture=priorfrac momfrac armassist raterisk2
                 height age priorfrac*age momfrac*armassist

/ctable pprob=(0.25 0.5) pevent=0.05 0.25;
run;
```

- Calculates Se, Sp, PPV and NPV for cutpoints 0.25 and 0.5
- 0.25 is the best cutpoint here
- 0.5 is the standard cutpoint and should always be tested

Determines PPV and NPV for prevalences 5% and 25%

Classification Table

Prob Event	Prob Level	Correct		Incorrect		Percentages				
		Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.050	0.250	80	252	123	45	67.0	64.0	67.2	90.7	2.7
0.050	0.500	19	352	23	106	89.9	15.2	93.9	88.5	4.5
0.250	0.250	80	252	123	45	66.4	64.0	67.2	60.6	15.2
0.250	0.500	19	352	23	106	74.2	15.2	93.9	54.8	23.1

Cutpoint

Prevalence of the outcome

PPV = 100% - False POS
NPV = 100% - False NEG

Issues

- Se, Sp and PPV are low
- A model tends to predict better in the data set it was based upon, i.e. it may perform (even) worse in other data sets
- But remember: Se and Sp are not indicators of goodness-of-fit

Se and Sp are not indicators of goodness-of-fit

Reason: Se and Sp depend on the proportion of $\hat{\pi}$s near the cutpoint (generally but not always 0.5)

Example

- Classify study subjects as follows:
  - if $\hat{\pi} < 0.5$, then the predicted outcome is 0
  - if $\hat{\pi} \geq 0.5$, then the predicted outcome is 1
- Assume perfect goodness-of-fit
- In other words, assume that the predicted values perfectly reflect the observed values

y=outcome
$\hat{\pi}$ > 0.5 but not by much

- If $\hat{\pi}$=0.51 for 100 subjects
  - 51 subjects have y=1
  - 49 subjects have y=0
- But $\hat{\pi}$>0.5 for 100 subjects
  - Predicted outcome=1 for all 100 subjects
  - 49 false positives

➔ Substantial decrease in Sp

y=outcome
$\hat{\pi}$ < 0.5 but not by much

- If $\hat{\pi}$=0.49 for 100 subjects
  - 49 subjects have y=1
  - 51 subjects have y=0
- But $\hat{\pi}$<0.5 for 100 subjects
  - Predicted outcome=0 for all 100 subjects
  - 49 false negatives

➔ Substantial decrease in Se

y=outcome
$\hat{\pi}$ > 0.5 by a lot

- If $\hat{\pi}$=0.95 for 100 subjects
  - 95 subjects have y=1
  - 5 subjects have y=0
- But $\hat{\pi}$>0.5 for 100 subjects
  - Predicted outcome=1 for all 100 subjects
  - 5 false positives

➔ Small decrease in Sp

y=outcome
$\hat{\pi}$ < 0.5 by a lot

- If $\hat{\pi}$=0.05 for 100 subjects
  - 5 subjects have y=1
  - 95 subjects have y=0
- But $\hat{\pi}$<0.5 for 100 subjects
  - Predicted outcome=0 for all 100 subjects
  - 5 false negatives

➔ Small decrease in Se

Example summary

- In this example, the goodness-of-fit of the model is the same in all situations
- But Se and Sp depend on $\hat{\pi}$ and are not the same in all situations

Area under the ROC curve

Measures the model's ability to discriminate between the diseased and the non-diseased

Background

- ROC stands for Receiver Operating Characteristic
- First used in signal detection theory: How well does the receiver detect a signal in the presence of noise?
- ROC curve = Plot of Se vs. 1-Sp
- Area under the ROC curve measures receiver's ability to discriminate between true and false signals

Area under the ROC curve - Idea

- Pair each diseased subject with each non-diseased subject and compare each pair's $\hat{\pi}$s
- Determine the proportion of pairs where $\hat{\pi}$ for the diseased subject is greater than $\hat{\pi}$ for the non-diseased subject
- This proportion is equal to the area under the ROC curve

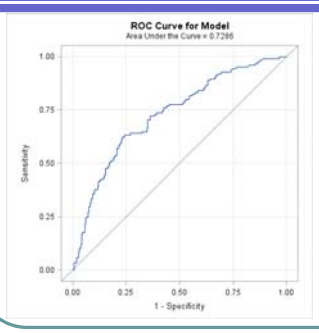
Area under the ROC curve

- 0.5 no discrimination (as good as coin toss)
- 0.7 - < 0.8 acceptable discrimination
- 0.8 - < 0.9 excellent discrimination
- ≥ 0.9 outstanding discrimination

Area under the ROC curve

- Close to quasi-complete separation would be required for the area under the ROC curve to be ≥ 0.9
- A diagonal line indicates no discrimination
- The faster Se increases with decreasing Sp (i.e., increasing 1-Sp), the better the discrimination

ROC curve for GLOW500 example



Area under the ROC curve = 0.729 ➔ acceptable discrimination