STATS 111/202

Lecture 12: More Model Comparisons

2/26/2025

Ana Maria Kenney

UC Irvine Department of Statistics

GLM Inference

4.2, 4.3, 4.4, 5.1

AIC

- The Akaike Information Criterion (AIC) provides a method for assessing the quality of your model through comparison of related models.

- It penalizes you for making the model more complicated (having to estimate more coefficients)

- Can think of this like an adjusted R-squared value to prevent the addition of irrelevant features

AIC Definition:

$$2k - 2\log(L_M)$$

Where k is the number of parameters to be estimated by the model
L_M Is the resulting maximized likelihood value

AIC

- This is based on the Kullback-Leibler divergence

- The KL divergence measures the information lost by using a proposed distribution (say g) instead of the true data generating distribution (say f)

- Now say we have two different proposed distributions, g_1 and g_2. We are comparing g_1 and g_2 by how close they are to the true distribution f

- Since we do not know the true f, Akaike showed that the KL divergence can still be estimated

- So, now we can estimate how much more or less information is lost by g_1 than by g_2

AIC

- AIC as a number, is not interpretable (1000 doesn’t mean very good or very bad necessarily)

- It is only useful when comparing two models to one another

- The **lower the AIC, the better**

- So we can now compare multiple models with different sets of covariates (they do not need to be nested) and use AIC to choose

ROC curve

- A measure of the predictive power of the binary (logistic) model we have uses what is known as the **ROC curve.**

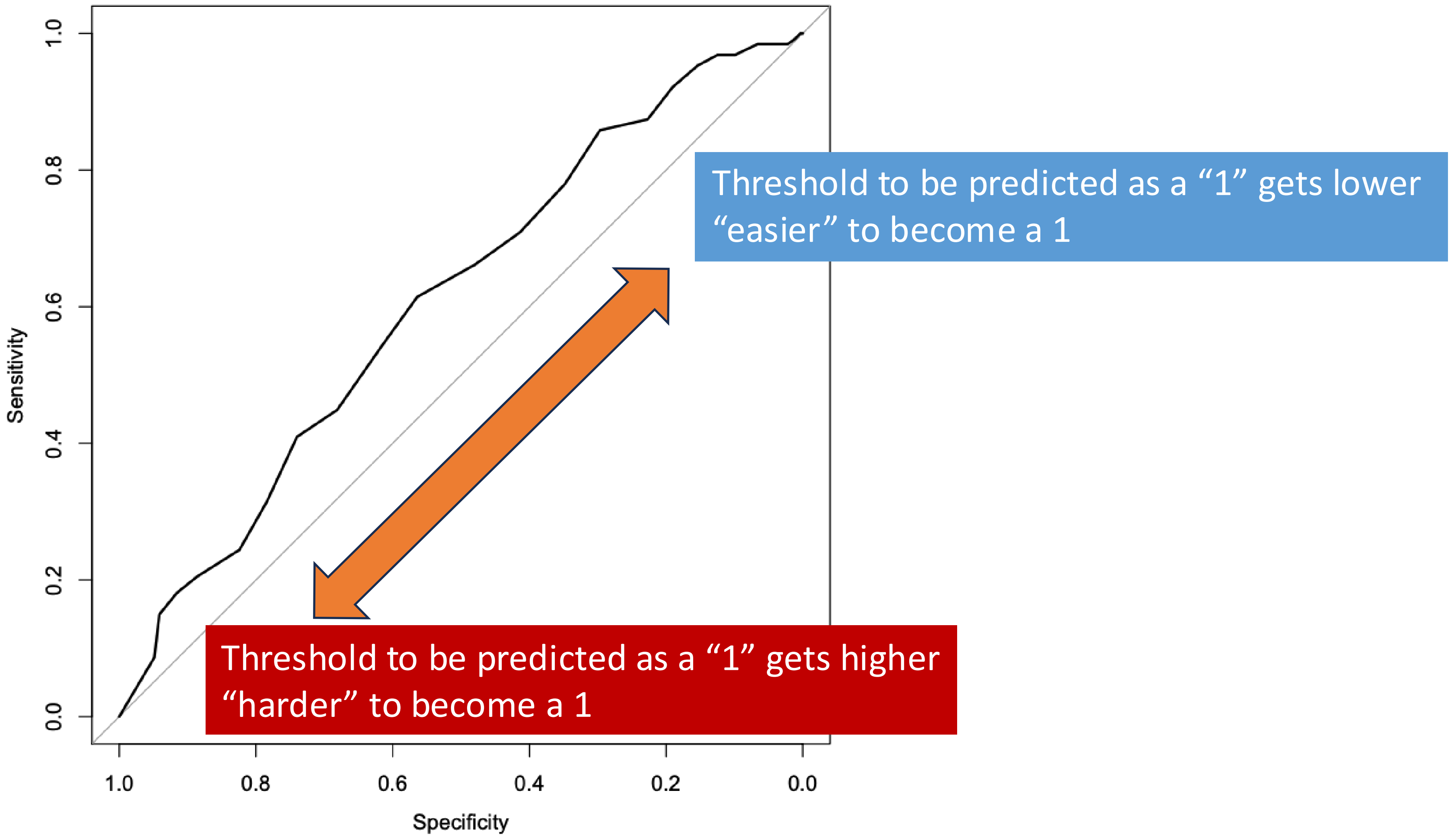
- A receiver operating characteristic curve, or ROC curve, is a plot that illustrates the performance of a binary classifier system **as its discriminant threshold is varied**

- With a logistic regression model, we only estimate the probability of an event, say Y = 1, given a certain set of explanatory variables X

- The ROC curve will assess the strength of predictions of the model as the threshold of what probability defines a prediction to be Y=1 is moved from 1 to 0

ROC curve

- The ROC curve has specificity on the X-axis and sensitivity on the Y-axis
- **Specificity:** the proportion of non-events (Y=0) that are correctly identified (true negatives)
- **Sensitivity:** the proportion of events (Y=1) that are correctly identified (true positives)
- Moving left to right on the graph, the probability threshold, say p, goes from 1 to 0
- All estimated probabilities that are **below p are designated $\hat{Y} = 0$** and all those **above p are designated $\hat{Y} = 1$**


Threshold to be predicted as a “1” gets lower
“easier” to become a 1
Threshold to be predicted as a “1” gets higher
“harder” to become a 1
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

ROC curve

- We care about the area under this curve. The maximum it can be is 1.
- We can judge model accuracy in terms of predictions by how close the area under the curve is to 1
- In R, we use the roc(Y~prob) function. This is in the pROC library
- Y is the vector of observations (1 or 0) and prob is the vector of the estimated probabilities of events obtained from the model
- Here, the area under the curve is 0.61, showing some strength in prediction but not necessarily very strong (this can heavily depend on problem context)

AIC and ROC Example