Checking Predictive Power

2/24/22

Response: $y = \begin{cases} 1 \\ 0 \end{cases}$ Prediction: $\hat{y} = \begin{cases} 1 \\ 0 \end{cases}$

Model gives us $\hat{\pi} \in [0,1]$.

Cut-off value: $\hat{y} = \begin{cases} 1 & \hat{\pi} > \pi_0 \\ 0 & \hat{\pi} \le \pi_0 \end{cases}$

For some chosen π_0. Common choices:

- $\pi_0 = 0.5$
- π_0 = Sample proportion of $y=1$

Classification Table:

		Predicted $\hat{y}=1$	Predicted $\hat{y}=0$
Observed	$y=1$	a	b
	$y=0$	c	d

Sensitivity: $P(\hat{y}=1 \mid y=1)$

Estimate: $\dfrac{a}{a+b}$

Specificity: $P(\hat{y}=0 \mid y=0)$

Estimate: $\dfrac{d}{c+d}$

Warning:
- Fit the model on the same data we're trying to predict

$\longrightarrow$ not fair assessment of predictive power.

Better - split data

1. Training $\longrightarrow$ fit model
2. Testing $\longrightarrow$ classification above

Limitations:

① Sensitive to our choice of π_0.

② Sensitive to the relative times that $y=1$ & $y=0$ in the data

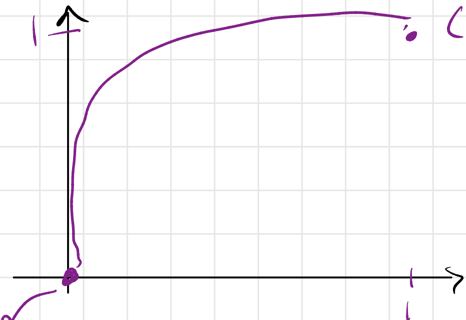
③ Loss of information classifying a probability $\hat{\pi}$ to 0, 1.

ROC — Receiver operating curves (characteristic)

— Calculate sensitivity & specificity for all $\pi_0 \in [0,1]$

$P(\hat{y}=1 \mid y=1)$

Sensitivity = true positive rate (want high)



$(1,1) \to \pi_0 = 0$

$\hat{y} = \begin{cases} 1 & \hat{\pi} > 0 \\ 0 & \hat{\pi} \le 0 \end{cases}$

$\Rightarrow \hat{y} = 1$

$\pi_0 = 1$

$\Rightarrow \hat{y} = \begin{cases} 1 & \hat{\pi} > 1 \\ 0 & \hat{\pi} \le 1 \end{cases}$

$\Rightarrow \hat{y} = 0$

1 − Specificity $= P(\hat{y}=1 \mid y=0)$

= false positive rate (want low)

C = "concordance index" = area under ROC curve

— estimates P(prediction & outcome are "concordant")

Higher c ⇒ more "concordant" = good predictive power

~ larger $\hat{\pi}$ → "larger" y

Framingham Example – Two predictors – sbp, sex

quantitative / binary

Fitted additive model:

$$\log\left(\frac{\hat\pi}{1-\hat\pi}\right) = -2.765 + 0.018\,x_1 - 0.783\,x_2$$

x_1 = systolic blood pressure (mmHg)

$$x_2 = \begin{cases} 1 & \text{female} \\ 0 & \text{male} \end{cases}$$

Interpretations in R code –

Males: $\log\left(\frac{\hat\pi}{1-\hat\pi}\right) = -2.765 + 0.018\,x_1$

Females: $\log\left(\frac{\hat\pi}{1-\hat\pi}\right) = -2.765 - 0.783 + 0.018\,x_1$

Females – Males : -0.783

Fitted interaction model:

$$\log\left(\frac{\hat\pi}{1-\hat\pi}\right) = -2.089 + 0.0128\,x_1 - 1.867\,x_2 + 0.0080\,x_1 x_2$$

Males $(x_2=0)$: $-2.089 + 0.0128\,x_1$

Females $(x_2=1)$: $(-2.089 - 1.867) + (0.0128 + 0.0080)\,x_1$

Females – Males : $-1.867 + 0.008\,x_1$

$$\log\left(\frac{\text{odds CHD Female}}{\text{odds CHD Male}} \,\middle|\, x_1\right)$$

How to interpret the interaction coef itself?

Multiplicative

= effect on an effect

= change in odds ratio

effect of sex on the effect of sbp on CHD

$$\frac{\text{odds of CHD sbp } x+1}{\text{odds of CHD sbp } x}$$

effect of sbp on the effect of sex on CHD

$$\frac{\text{odds CHD Female}}{\text{odds CHD Males}}$$