

Đánh giá mô hình hồi qui logistic

Tuan V. Nguyen

Senior Principal Research Fellow, Garvan Institute of Medical Research
Professor, UNSW School of Public Health and Community Medicine
Professor of Predictive Medicine, University of Technology Sydney
Adj. Professor of Epidemiology and Biostatistics,
School of Medicine Sydney, University of Notre Dame Australia



Ba tiêu chí cho một mô hình tiên lượng

- **Discrimination** – phân định
- **Calibration** – chính xác
- **Reclassification** – tái phân nhóm

Discrimination – phân định

Hai thước đo độ tin cậy của một mô hình

- Sensitivity – độ nhạy
- Specificity – độ đặc hiệu

Độ nhạy

- Trong số những người mắc bệnh, bao nhiêu % có tiên lượng dương tính?
- Gold standard – mắc bệnh trong thực tế
- Test result – mô hình tiên lượng

Độ đặc hiệu

- Trong số những người **không** mắc bệnh, bao nhiêu % có tiên lượng *âm tính*
- Gold standard – mắc bệnh trong thực tế
- Test result – mô hình tiên lượng

Ví dụ

Tiên lượng	Bệnh (n=177)	Không bệnh (n=81)
+ve (trên 20%)	120	20
-ve (dưới 20%)	57	61
Tổng số	177	81

Ví dụ

Tiên lượng	Bệnh (n=177)	Không bệnh (n=81)
+ve (trên 20%)	120	20
-ve (dưới 20%)	57	61
Tổng số	177	81

$$\text{Sensitivity} = 120 / 177 = 0.68$$

$$\text{Specificity} = 61 / 81 = 0.75$$

Tỉ lệ dương tính giả (false +ve)

Tiên lượng	Bệnh (n=177)	Không bệnh (n=81)
+ve (trên 20%)	120	20
-ve (dưới 20%)	57	61
Tổng số	177	81

Sensitivity (dương tính thật) = $120 / 177 = 0.68$

Specificity = $61 / 81 = 0.75$

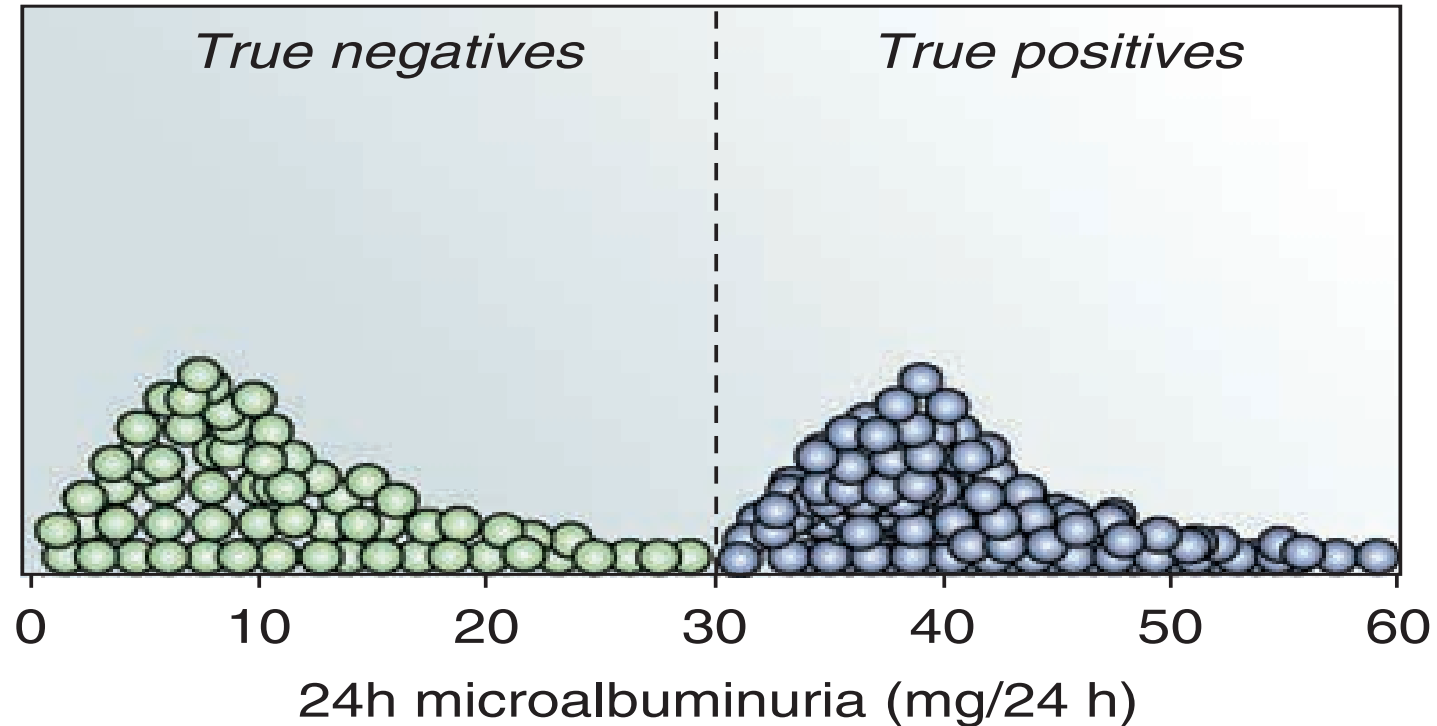
Dương tính giả = $1 - 0.75 = 0.25$

ROC curve

- Receiver operating characteristic (ROC) curve
- Đo lường khả năng phân định (**power of discrimination**) của một xét nghiệm hay mô hình tiên lượng
- Thường dùng cho các kết quả biến liên tục
- Y-axis (trục tung): true positive (sensitivity)
- X-axis (trục hoành): false positive (1-specificity)

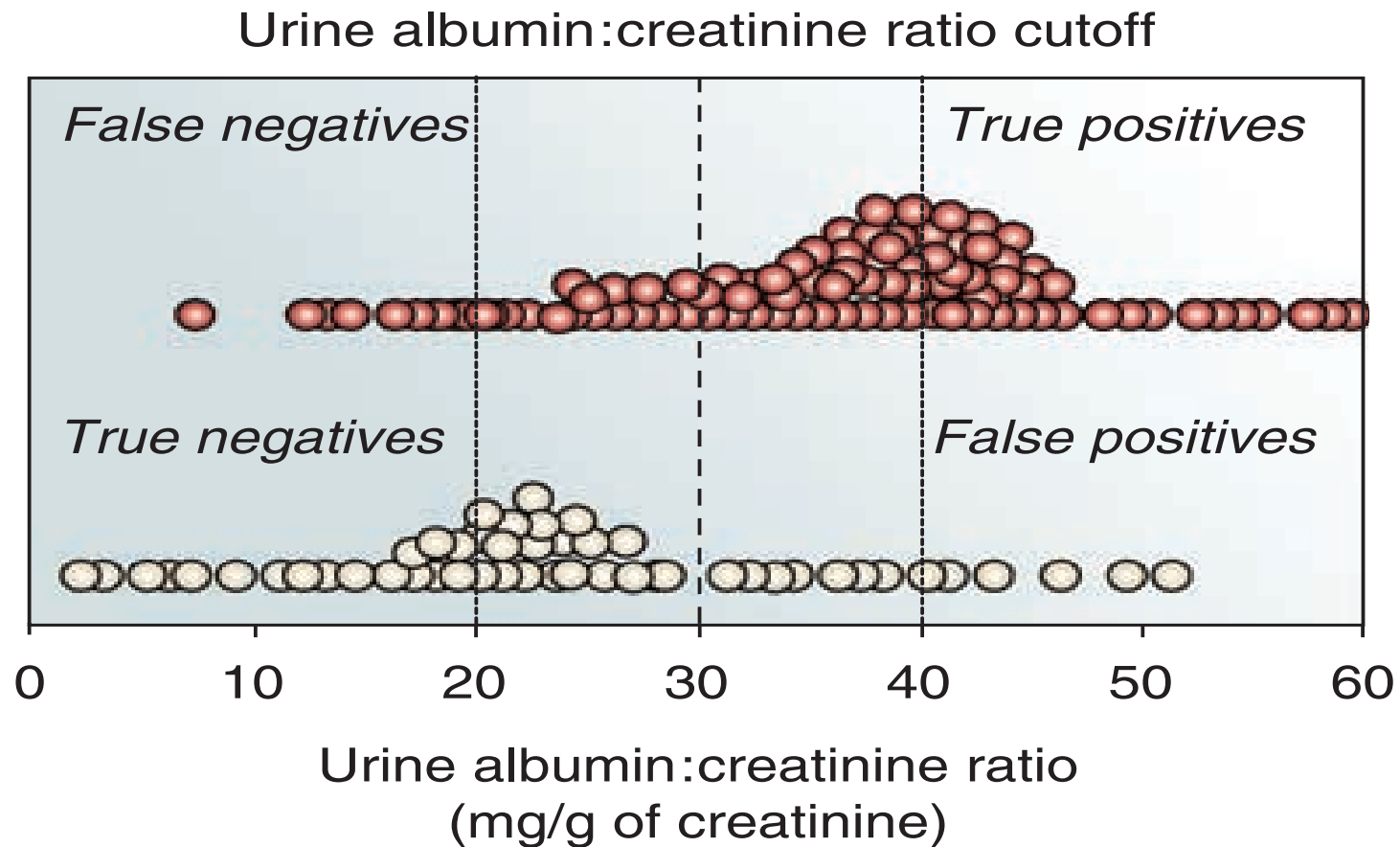
Một ví dụ về một xét nghiệm hoàn hảo

24 h microalbuminuria cutoff



- Patients with renal dysfunction
- Healthy subjects

Nhưng trong thực tế

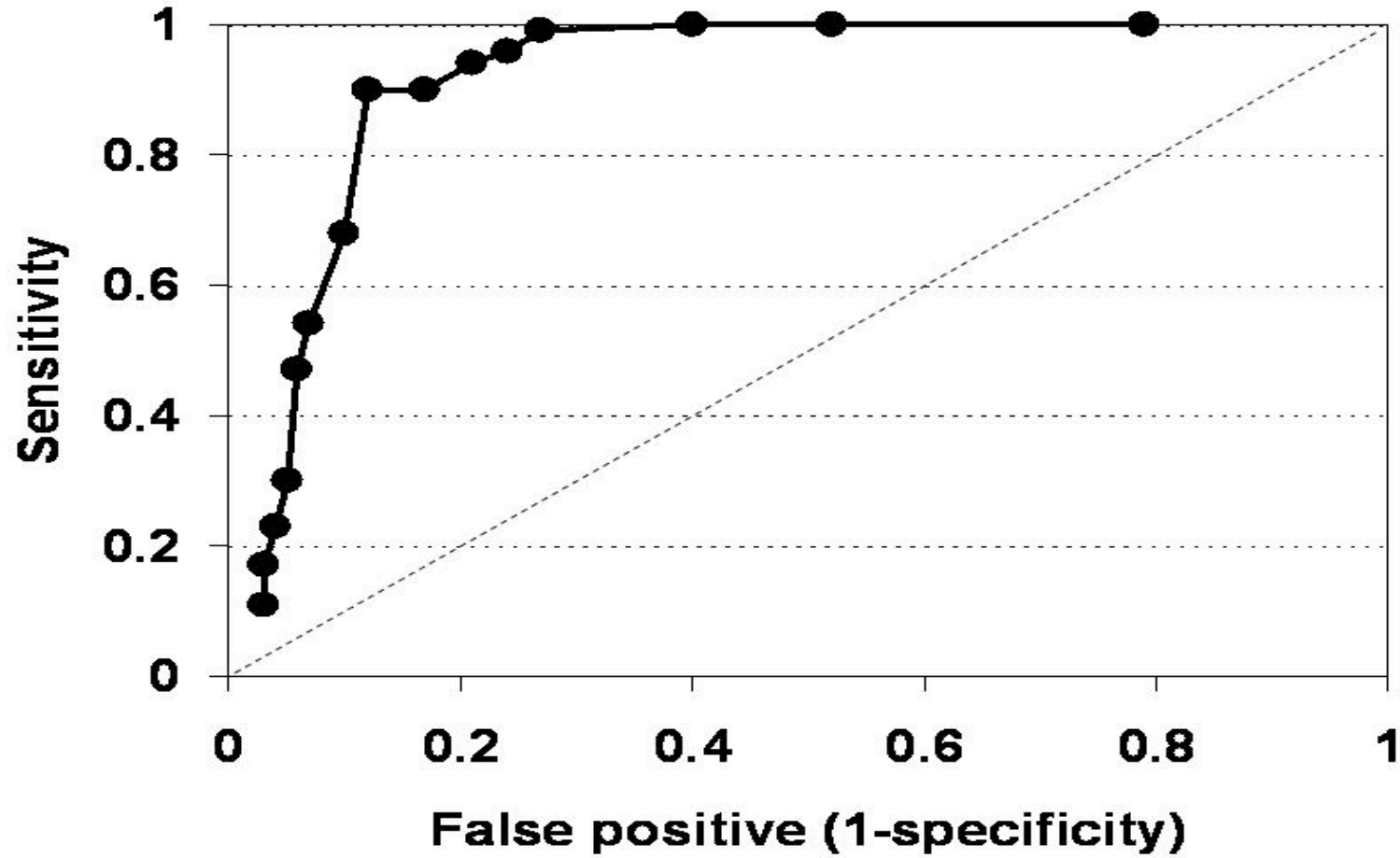


- Patients with renal dysfunction
- Healthy subjects

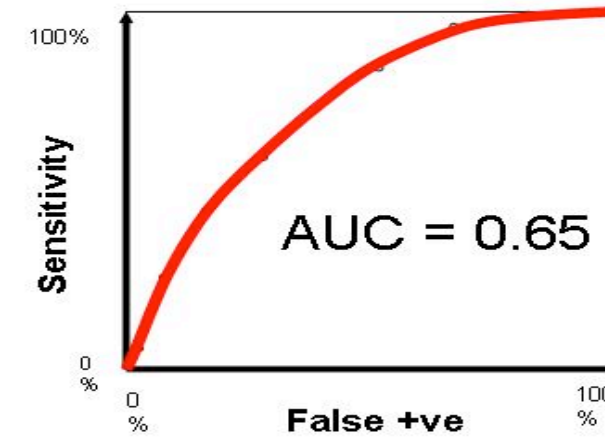
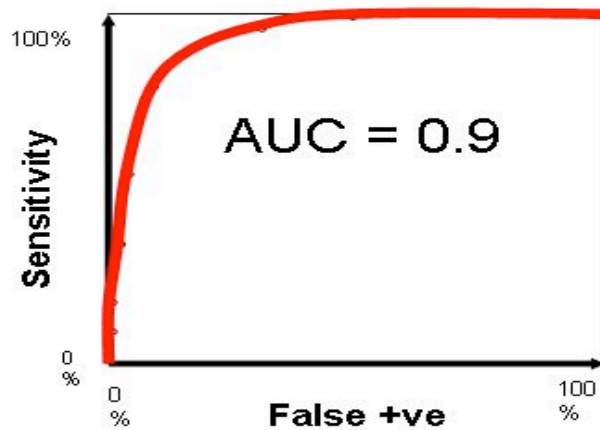
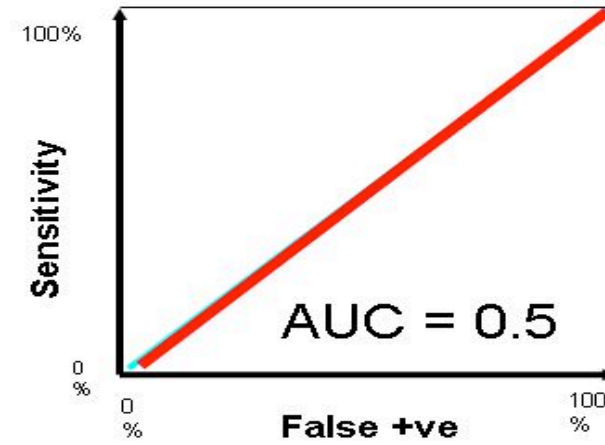
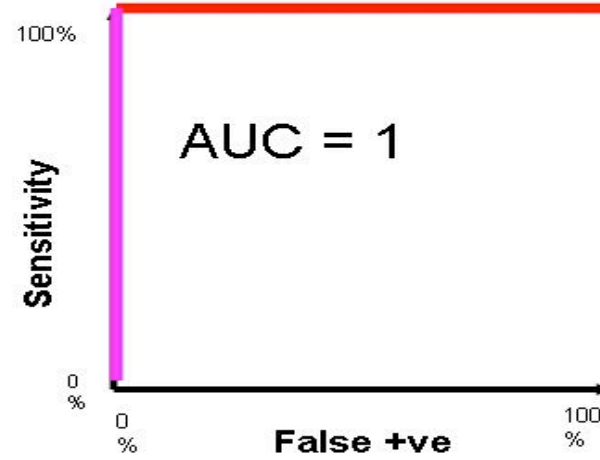
Ví dụ về tiên lượng ung thư tiền liệt tuyến

PSA level (ng/mL)	Sensitivity	Specificity
≥ 1	1.00	0.21
≥ 2	1.00	0.48
≥ 3	1.00	0.60
≥ 4	0.99	0.73
≥ 5	0.96	0.76
≥ 6	0.94	0.79
≥ 7	0.90	0.83
≥ 8	0.90	0.88
≥ 9	0.68	0.90
≥ 10	0.54	0.93
≥ 11	0.47	0.94
≥ 12	0.30	0.95
≥ 13	0.23	0.96
≥ 14	0.17	0.97
≥ 15	0.11	0.97

ROC curve



Diện tích dưới đường biểu diễn (area under the curve - AUC)



Diễn giải AUC

AUC	Meaning
>0.90	Excellent test
0.80 to 0.90	Good
0.70 to 0.80	Fair
0.60 to 0.70	Poor
0.50 to 0.60	Fail

Ý nghĩa thật của AUC

- Định nghĩa: *Probability that a randomly selected pair of healthy individual and patient, the test result will be higher in the patient than in the healthy individual* (**xác suất mà một cặp bệnh nhân và người bình thường được chọn, và bệnh nhân có giá trị tiên lượng cao hơn người bình thường**)
- Khó hiểu!

Một ví dụ đơn giản

- Chúng ta có 7 người được theo dõi 5 năm
- 4 người mắc bệnh ung thư tiền liệt tuyến, và giá trị PSA là:

8, 2, 6, 3

- 3 người không mắc bệnh, giá trị PSA:

3, 2, 6

Tổ hợp

- 4 giá trị PSA của 4 bệnh nhân
- 3 giá trị PSA của 3 người không bệnh
- Tổng số cặp có thể: $4 \times 3 = 12$
- $AUC = \text{số cặp mà PSA bệnh nhân cao hơn PSA của người không mắc bệnh} \textit{ chia cho } 12.$

Nếu chúng bắt cặp bệnh nhân 1 với 3 người trong nhóm không bệnh:

Bệnh	Không bệnh	Chú ý
8	3	Concordant ($8 > 3$)
2	2	Concordant ($8 > 2$)
6	6	Concordant ($8 > 6$)
3		



Nếu chúng bắt cặp bệnh nhân 1 với 3 người trong nhóm không bệnh:

Bệnh	Không bệnh	Chú ý
8	3	Concordant ($8 > 3$)
2	2	Concordant ($8 > 2$)
6	6	Concordant ($8 > 6$)
3		

Đến bệnh nhân thứ 2

Bệnh	Không bệnh	Chú ý
8	3	Discordant ($2 < 3$)
2	2	Tie ($2 = 2$)
6	6	Discordant ($2 < 6$)
3		


Đến bệnh nhân thứ 3

Bệnh	Không bệnh	Chú ý
8	3	Concordant ($6 > 3$)
2	2	Concordant ($6 > 2$)
6	6	Tie ($6 = 6$)
3		



và bệnh nhân thứ 4

Bệnh	Không bệnh	Chú ý
8	3	Tie ($3 = 3$)
2	2	Concordant ($3 > 2$)
6	6	Discordant ($3 < 6$)
3		



AUC (hay c-index)

- $AUC = (\text{concordant} + 0.5 \text{ ties}) / \text{tổng số cặp}$
- Chúng ta có:
 - 6 cặp concordance
 - 3 cặp discordance
 - 3 cặp ties (trùng)
 - $AUC = (6 + 1.5) / 12 = 0.625$

Ví dụ: tính AUC cho mô hình hồi qui logistic

- Mục tiêu: tiên lượng thu nhập trên \$50K
- Yếu tố: Age, Education.Years, Occupation, Race, Sex
- Câu hỏi: 4 yếu tố này tiên lượng thu nhập chính xác như thế nào?
- Trả lời: AUC

Package "pROC"

- pROC package
(<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-77>)
- Có thể dùng để tính AUC và vẽ biểu đồ ROC
- Các bước cần thiết
 - Xây dựng mô hình m
 - Tính giá trị tiên lượng dựa trên tham số của m
 - Tính AUC

Đọc dữ liệu income

```
inc = read.csv("~/Dropbox/_Conferences and Workshops/TDTU  
2018/Datasets/adult.data.csv")
```

Fit mô hình hồi qui logistic

```
m = glm(Income ~ Age + Education.Years + Occupation + Race + Sex,  
family=binomial, data=inc)
```

Dùng pROC

```
library(pROC)
```

Tính xác suất tiên lượng

```
pred = predict(m, type="response")
```

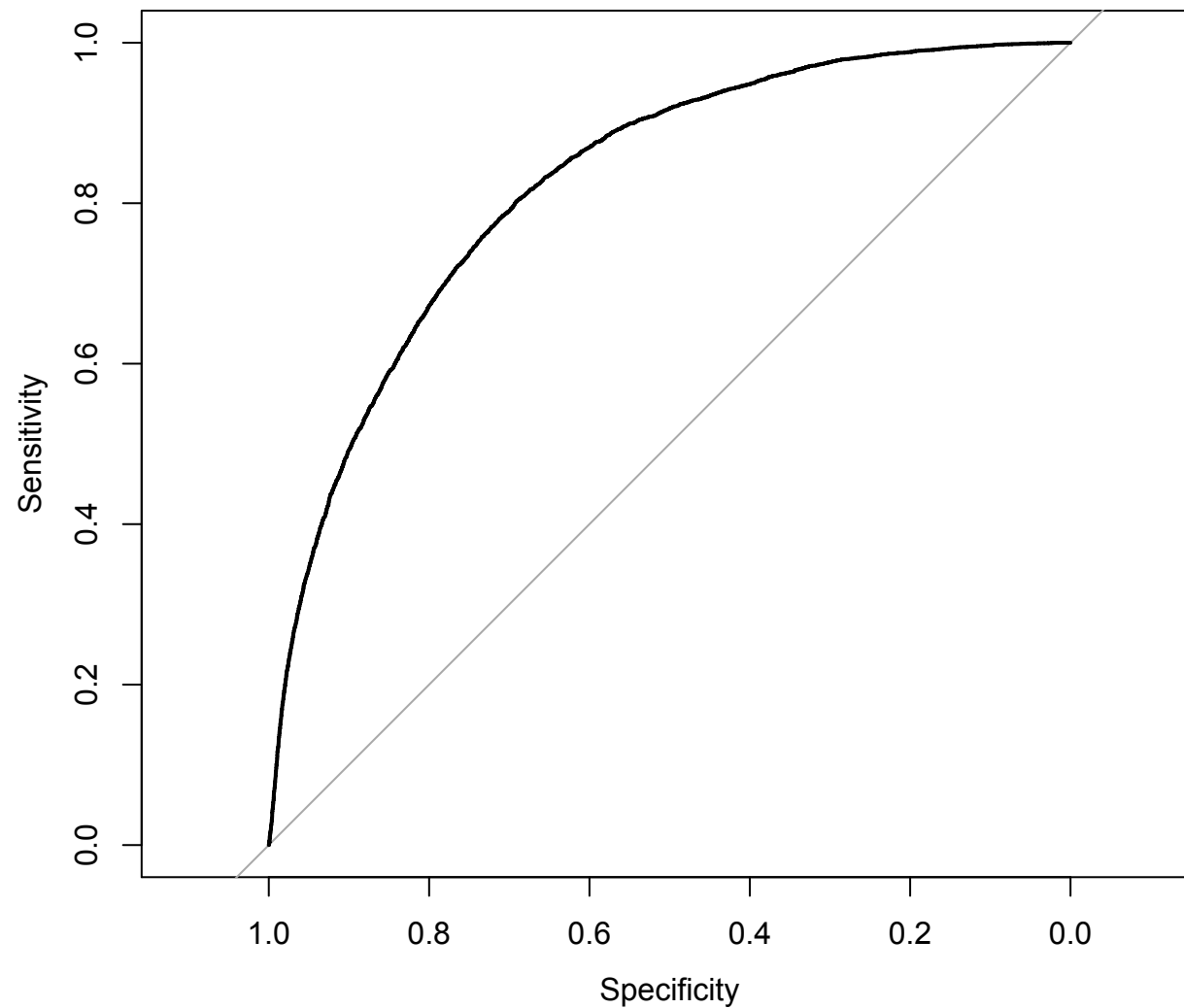
Tạo ra object roc, tính AUC

```
roccurve = roc(inc$Income ~ pred)  
auc(roccurve)  
ci(roccurve)  
plot(roccurve)  
ci.thresholds(roccurve)
```

```
> auc(roccurve)
Area under the curve: 0.8241

> ci(roccurve)
95% CI: 0.819-0.8291 (DeLong)

> plot.roc(roccurve)
```



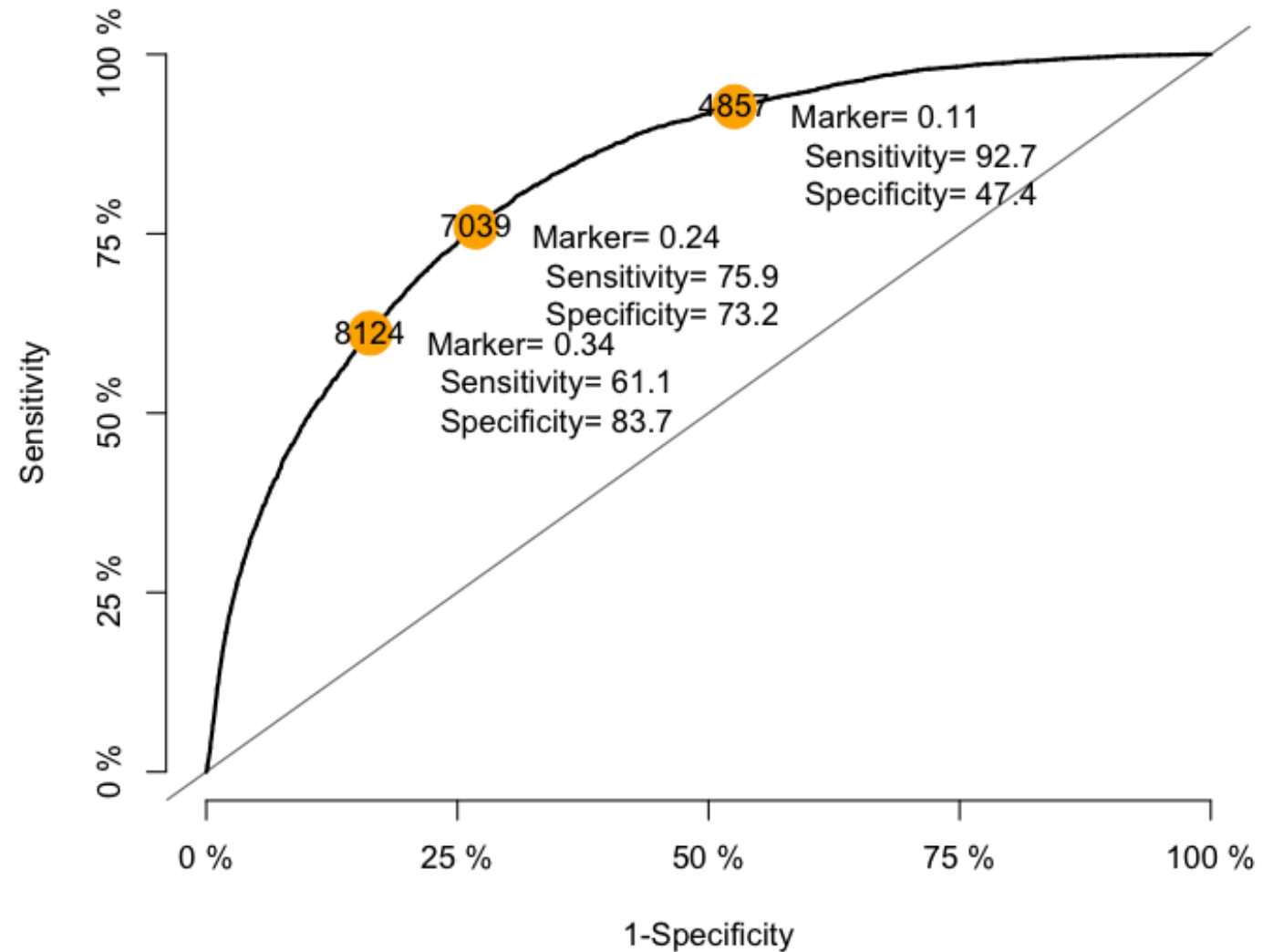
Dùng ModelGood để tìm giá trị
"tối ưu"

```
library(ModelGood)
```

```
r = Roc(Inc50 ~ Age +  
Education.Years + Occupation  
+ Race + Sex, data=inc)
```

```
plot(r)
```

```
click.Roc(r)
```



Calibration – chính xác

Đánh giá độ chính xác - calibration

- AUC không phản ánh mô hình tiên lượng có **chính xác** hay không
- Chính xác: giá trị tiên lượng gần với giá trị thực tế
- **Calibration** = phản ánh độ chính xác

Residual – phần dư

- Residual là hiệu số của
 - tình trạng bệnh của một cá nhân, và
 - xác suất tiên lượng (predicted probability) do mô hình tiên lượng

$$\text{Residual (cá nhân } i) = Y_i - P_i$$

Residual của mô hình hồi qui logistic

Bệnh nhân i	Y_i	P_i	Residual
1	0	2.31	-2.31
2	0	1.91	-1.91
3	1	98.11	1.89
4	1	79.58	20.42
5	0	4.21	-4.21
6	1	98.81	1.19
7	1	64.72	35.28
.	.	.	.
.	.	.	.

Chỉ số Brier

- Brier score = residual bình phương

$$B = \frac{1}{N} \sum (Y_i - P_i)^2$$

- Brier score càng thấp càng tốt

Residual của mô hình hồi qui logistic

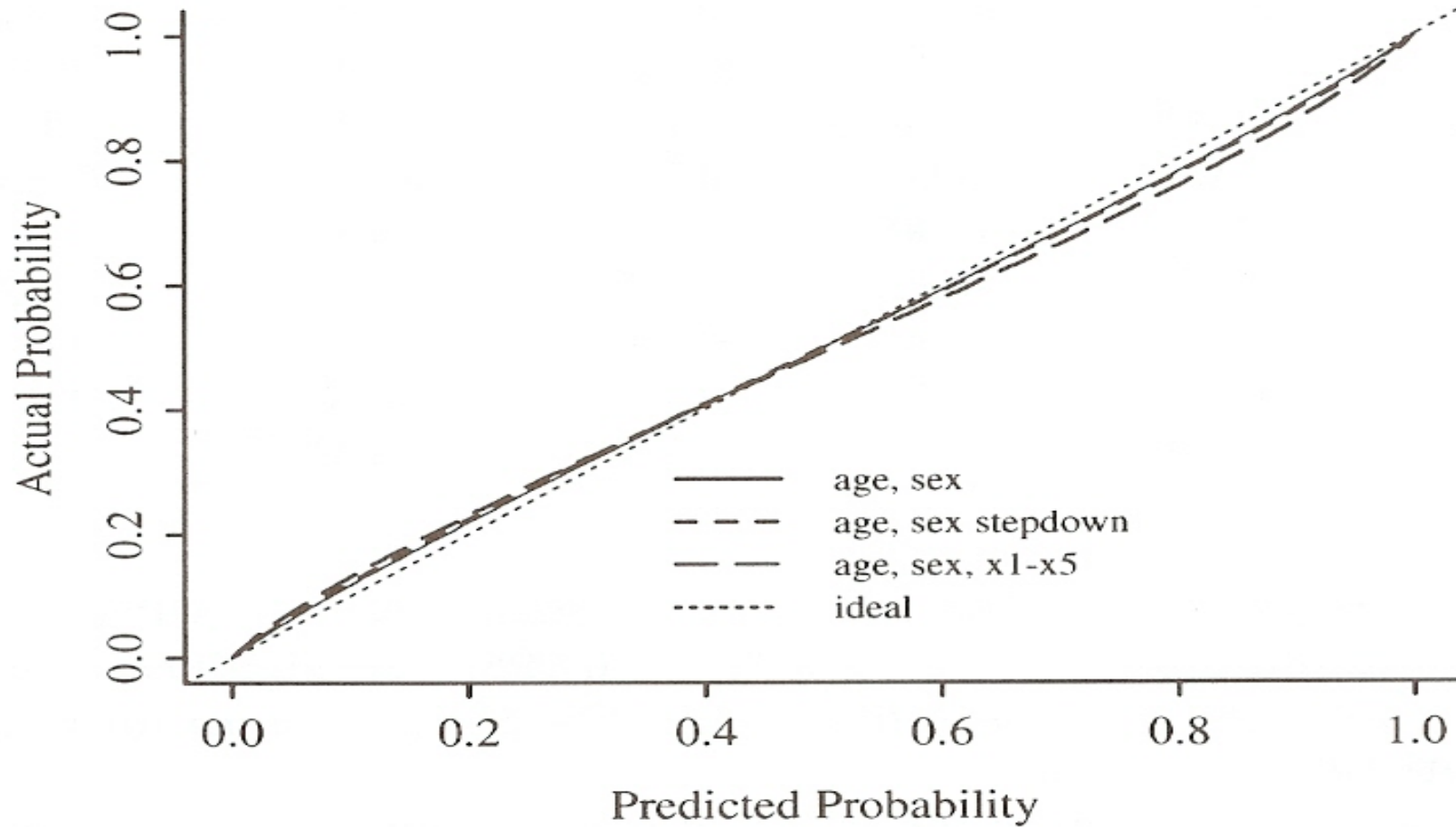
Cá nhân i	Y_i	P_i	Residual ($Y_i - P_i$)	$(Y_i - P_i)^2$
1	0	2.31	-2.31	5.34
2	0	1.91	-1.91	3.65
3	1	98.11	1.89	3.57
4	1	79.58	20.42	416.98
5	0	4.21	-4.21	17.72
6	1	98.81	1.19	1.42
7	1	64.72	35.28	1244.68
.
.
Total				S

$$\text{Brier Score} = S / N$$

Một cách thể hiện calibration

Nhóm nguy cơ (xác suất)	Số cá nhân trong thực tế	Số cá nhân tiên lượng
1 (0.00 – 0.20)	2	5
2 (0.21 – 0.40)	5	6
3 (0.41 – 0.60)	10	8
4 (0.61 – 0.80)	16	17
5 (0.81 – 1.00)	52	51

Calibration Curve



Harrell 2001

Đọc dữ liệu income

```
inc = read.csv("~/Dropbox/_Conferences and Workshops/TDTU  
2018/Datasets/adult.data.csv")
```

Fit mô hình hồi qui logistic

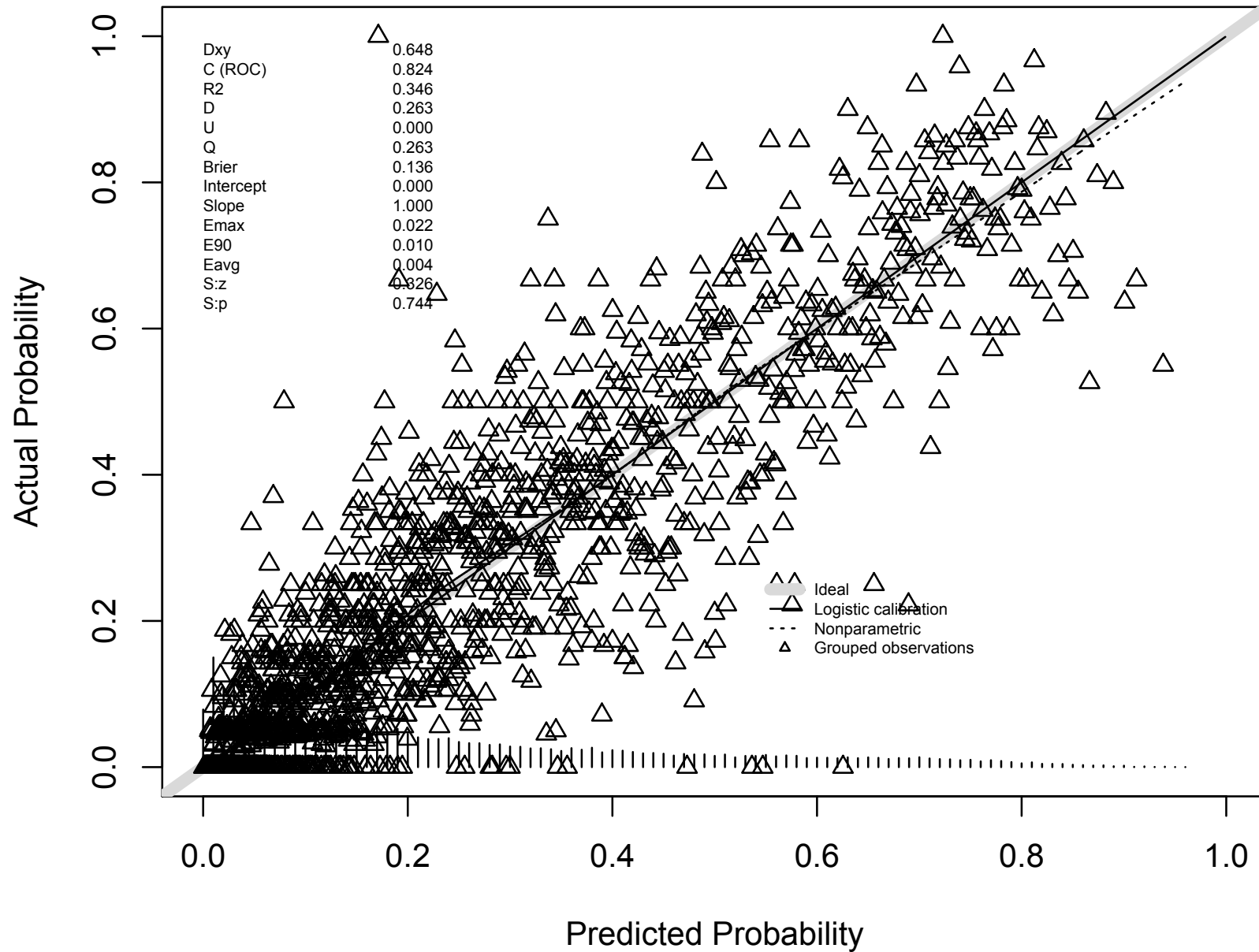
```
f = lrm(Inc50 ~ Age + Education.Years + Occupation + Race + Sex,  
data=inc)  
f
```

Tính xác suất tiên lượng

```
pred.logit = predict(f)  
phat = 1/(1+exp(-pred.logit))
```

Calibration

```
val.prob(phat, inc$Inc50, m=20, cex=.5)
```



> f

Logistic Regression Model

```
lrm(formula = Income ~ Age + Education.Years + Occupation + Race +  
     Sex, data = inc)
```

		Model Likelihood Ratio Test		Discrimination Indexes		Rank Discrim. Indexes	
Obs	32561	LR chi2	8569.01	R2	0.346	C	0.824
<=50K	24720	d.f.	21	g	1.754	Dxy	0.648
>50K	7841	Pr(> chi2)	<0.0001	gr	5.775	gamma	0.649
max deriv	3e-05			gp	0.236	tau-a	0.237
				Brier	0.136		

Dùng package "givitiR"

```
f = lrm(Inc50 ~ Age +  
Education.Years + Occupation +  
Race + Sex, data=inc)
```

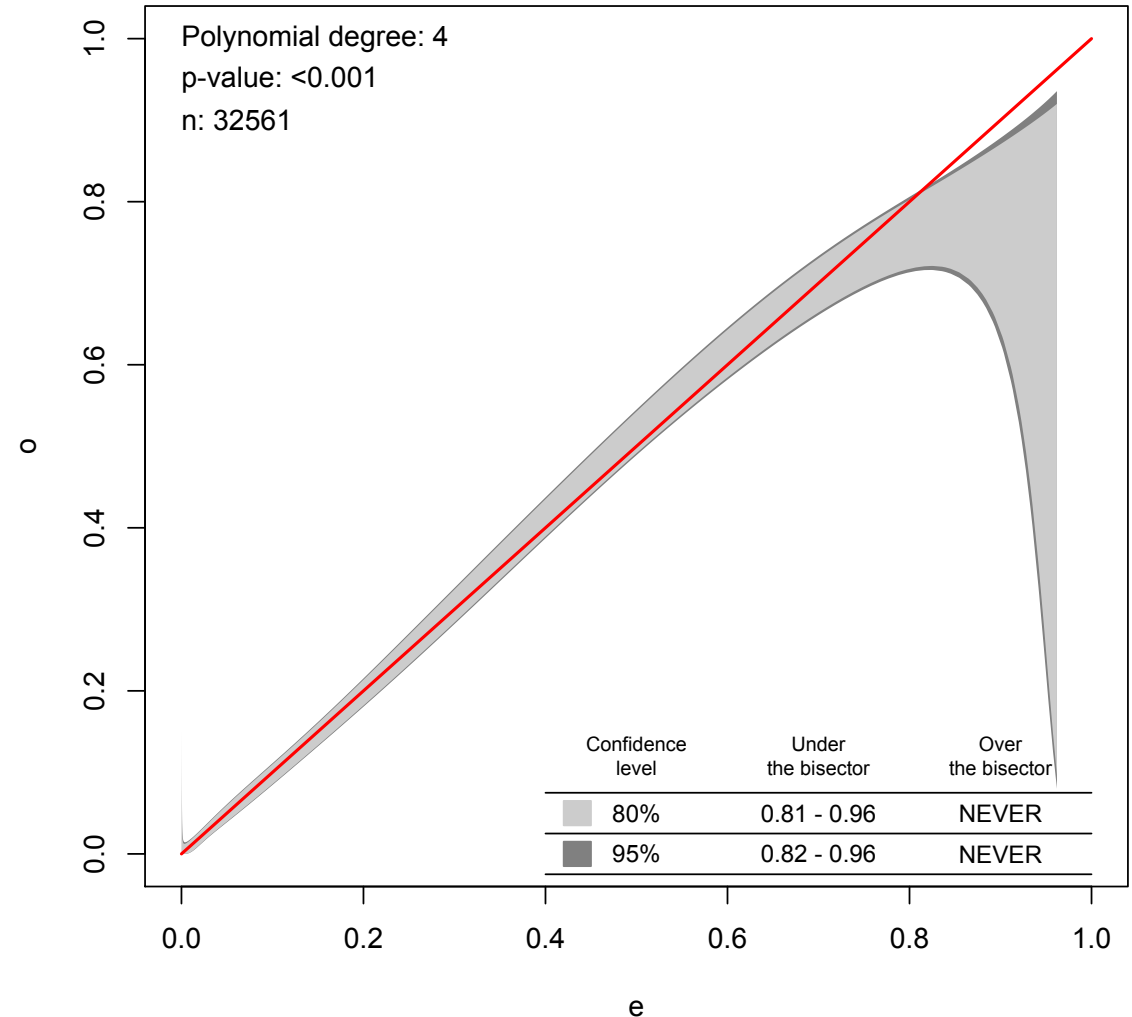
```
pred.logit = predict(f)
```

```
phat = 1/(1+exp(-pred.logit))
```

```
library(givitiR)
```

```
cb = givitiCalibrationBelt(o =  
inc$Inc50, e=phat, delev =  
"external")  
plot(cb)
```

GiViTI Calibration Belt



Đọc dữ liệu income

```
inc = read.csv("~/Dropbox/_Conferences and Workshops/TDTU  
2018/Datasets/adult.data.csv")
```

Fit mô hình hồi qui logistic

```
f = lrm(Inc50 ~ Age + Education.Years + Occupation + Race + Sex,  
data=inc, subset=1:10000)
```

Tính xác suất tiên lượng

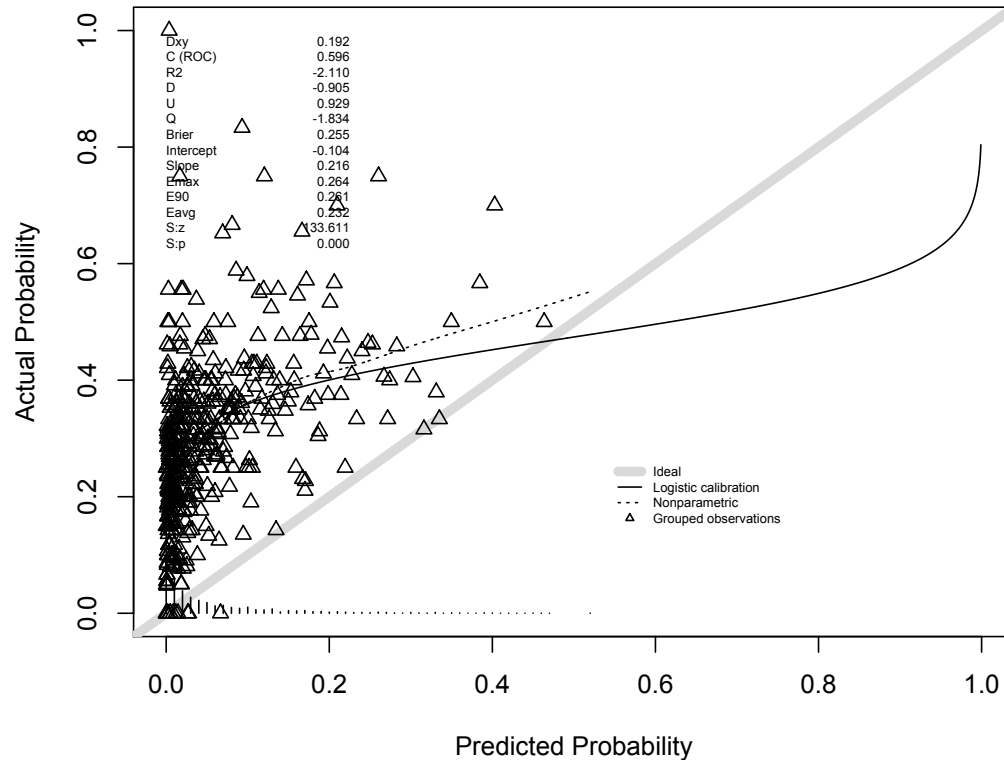
```
pred.logit = predict(f, inc[1:10000,])  
phat = 1/(1+exp(-pred.logit))
```

Calibration

```
val.prob(phat, inc$Inc50[10001:20000], m=20, cex=.5)
```

```
> val.prob(phat, inc$Inc50[10001:20000], m=20, cex=.5)
```

Dxy	C (ROC)	R2	D	D:Chi-sq
0.1919587	0.5959793	-2.1099648	-0.9051756	-9050.7558517
D:p	U	U:Chi-sq	U:p	Q
NA	0.9289382	9291.3817905	0.0000000	-1.8341138
Brier	Intercept	Slope	E _{max}	E ₉₀
0.2549178	-0.1039013	0.2164229	0.2640170	0.2611439
E _{avg}	S:z	S:p		
0.2323711	133.6106877	0.0000000		

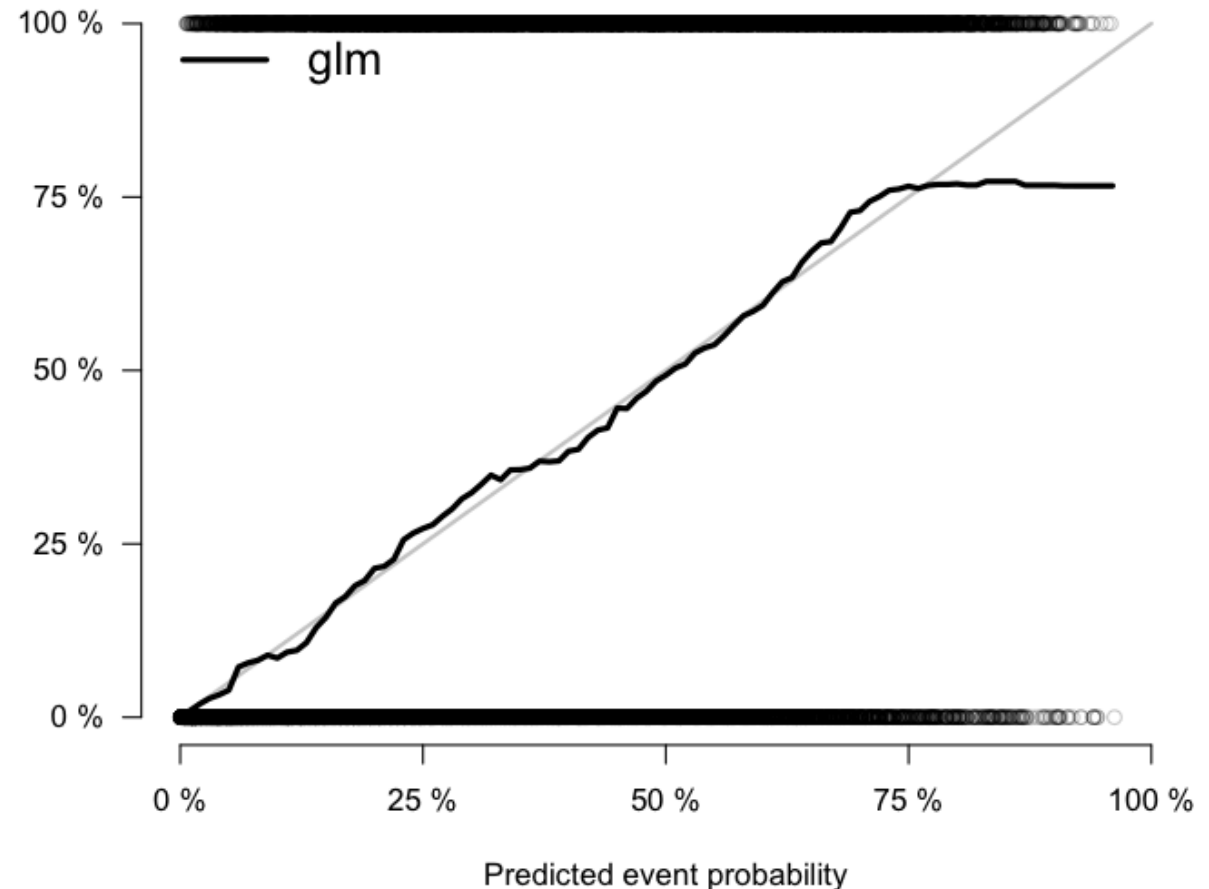


Dùng package "ModelGood"

```
library(ModelGood)
```

```
m = glm(Inc50 ~ Age + Education.Years + Occupation + Race + Sex,  
data=inc, family=binomial)
```

```
calPlot2(m)
```



Đánh giá mô hình tiên lượng: Tóm lược

- Hai tiêu chí để đánh giá một mô hình tiên lượng hồi qui logistic: discrimination và calibration
- Discrimination: dùng AUC và ROC
- Calibration (goodness-of-fit): Brier score và giá trị "predicted-observed"