

Raport

Klaudia Balcer

17 marca 2020

Spis treści

1	model z interakcją a model bez interakcji	2
2	modele z różnymi funkcjami linkującymi	2
2.1	logit	3
2.2	probit	3
2.3	cauchit	4
2.4	cloglog	4

1 model z interakcją a model bez interakcji

model	untermcept - pvalue	numeracy - pvalue	anxiety - pvalue	anxiety*numeracy - pvalue	AIC
z interakcją	0.985	0.684	0.891	0.774	36.201
bez interakcji	0.03623	0.01995	0.00396	—	34.286

Patrzając na wyniki testów istotności w modelu z interakcją, dochodzimy do wniosków, że najprawdopodobniej, żadna ze zmiennych objaśniających nie ma istotnego wpływu na zmienną objaśnianą. Nie jest to dobry model.

Patrzając na p-wartości w modelu bez interakcji, widzimy, że wszystkie uzyskane współczynniki są istotne - istnieją korelacje między obiema zmiennymi objaśniającymi a zmienną objaśnianą. Również AIC jest mniejsze w modelu bez interakcji, co również przemawia na jego korzyść.

2 modele z różnymi funkcjami linkującymi

Dla wszystkich funkcji linkujących krzywe ROC mają dość dużą całość, co wskazuje, że mogą to być dobre modele. Przyjrzyjmy się im bliżej, by móc przedstawić subtelniejsze różnice.

Patrzając na testy istotności, widzimy, że model z użyciem funkcji linkującej cauchit nie wskazuje na korelację zmiennych objaśniających i objaśnianej. Ma on też wyraźnie większą wartość AIC. Użycie funkcji linkującej cauchit zdecydowanie nie jest optymalne.

Wątpliwości może budzić również użycie funkcji cloglog. Zauważamy w tym modelu silną korelację między zmiennymi X_i i Y_i . Testy jednak odrzucają istotność współczynnika b_0 . Nie jest to najlepszy model, ale istotność współczynników dla zmiennych objaśniających jest cenna i skłania do dalszych badań nad danymi.

Patrzając na współczynnik AIC jako najlepszą funkcję linkującą uznajemy probit. Wskazuje on na istotność wszystkich współczynników modelu, co razem z dużym polem pod krzywą ROC zdecydowanie przemawia, za odpowiednością tego modelu.

Podane wyżej wartości predykcyjnych prawdopodobieństw obliczyłam za pomocą funkcji `predict` z parametrem `type = "response"`. Można to było również obliczyć *ręcznie* ze wzoru:

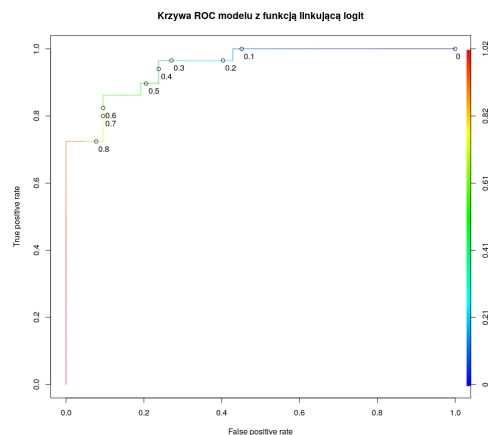
$$p_i = \frac{\exp\{b_0 + b_1 X_{1,i} + \dots + b_{p-1} X_{p-1,i}\}}{1 + \exp\{b_0 + b_1 X_{1,i} + \dots + b_{p-1} X_{p-1,i}\}}$$

```
model <- glm(success~numeracy+anxiety, data=A, family=binomial(link = "cloglog"))
new <- data.frame(numeracy = c(10), anxiety = c(13))
predict.glm(model, new, type = "response")
b <- model$coefficients
exp(b[1]+b[2]*10+b[3]*13)/(1+exp(b[1]+b[2]*10+b[3]*13))
```

To działanie dało identyczne wyniki (czego też się spodziewaliśmy), toteż nie odnotowywałam tego w sprawozdaniu dla poszczególnych grup.

Do generowania krzywych ROC użyłam biblioteki ROCR.

2.1 logit

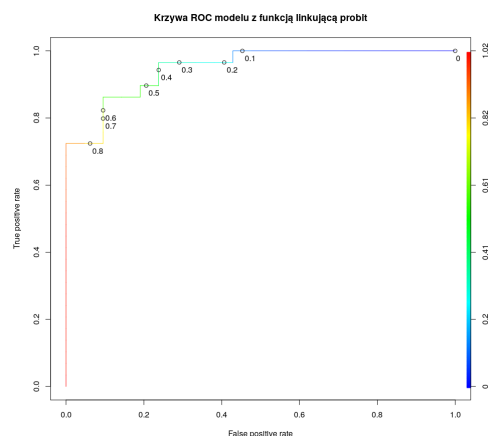


	intercept	numeracy	anxiety
wartość estymatora	14.2386	0.5774	-1.3841
wartość statystyki	2.094	2.327	-2.881
p-wartość	0.03623	0.01995	0.00396
wynik tetstu istotności	+	+	++

$$AIC = 34.286$$

$$\hat{P}(n = 10, a = 13) = 0.8827987$$

2.2 probit

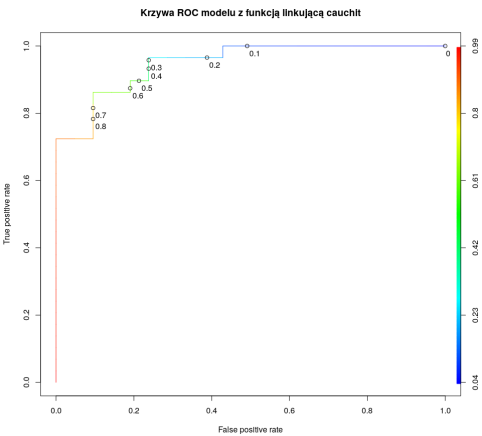


	intercept	numeracy	anxiety
wartość estymatora	8.2573	0.3371	-0.8039
wartość statystyki	2.247	2.464	-3.191
p-wartość	0.02466	0.01374	0.00142
wynik tetstu istotności	+	+	++

$$AIC = 33.854$$

$$\hat{P}(n = 10, a = 13) = 0.8806045$$

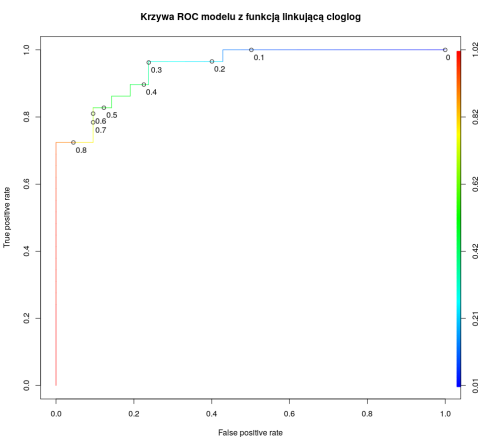
2.3 cauchit



	intercept	numeracy	anxiety
wartość estymatora	18.3830	0.7323	-1.7741
wartość statystyki	1.495	1.545	-1.789
p-wartość	0.1350	0.1224	0.0735
wynik tetstu istotności	-	-	.

$AIC = 37.115$
 $\hat{P}(n = 10, a = 13) = 0.8848509$

2.4 cloglog



	intercept	numeracy	anxiety
wartość estymatora	9.0006	0.4024	-0.9390
wartość statystyki	1.935	2.644	-2.827
p-wartość	0.05304	0.00819	0.00470
wynik tetstu istotności	.	++	++

$AIC = 34$
 $\hat{P}(n = 10, a = 13) = 0.8963072$