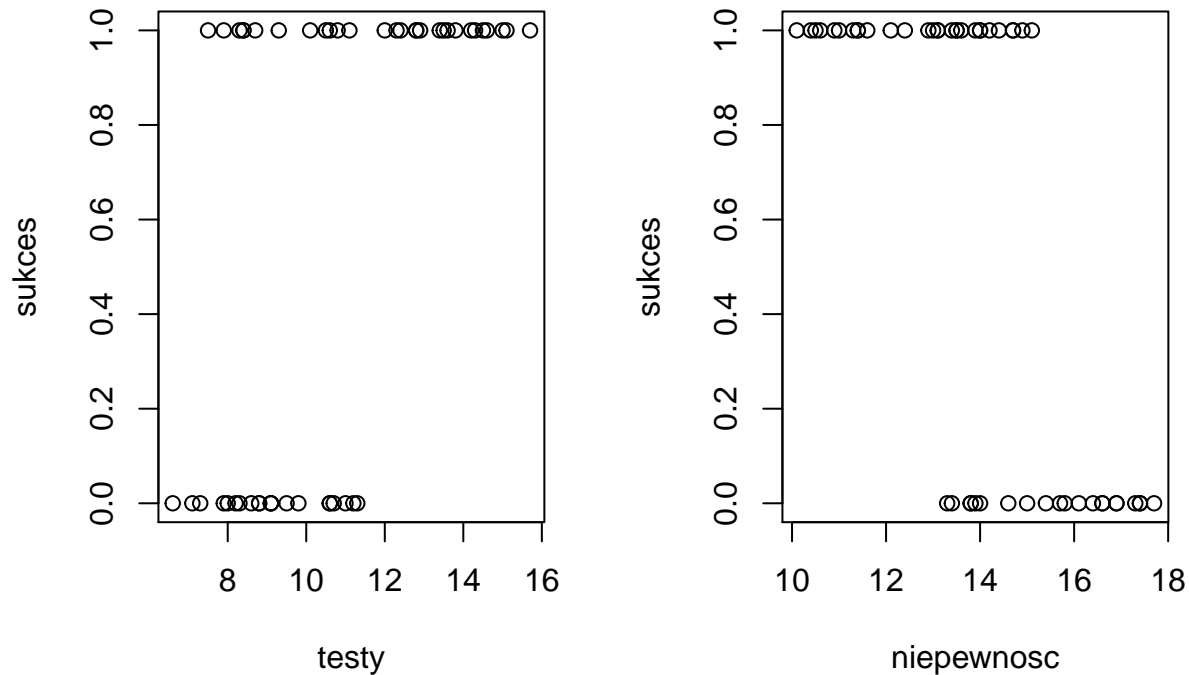


Lista 1

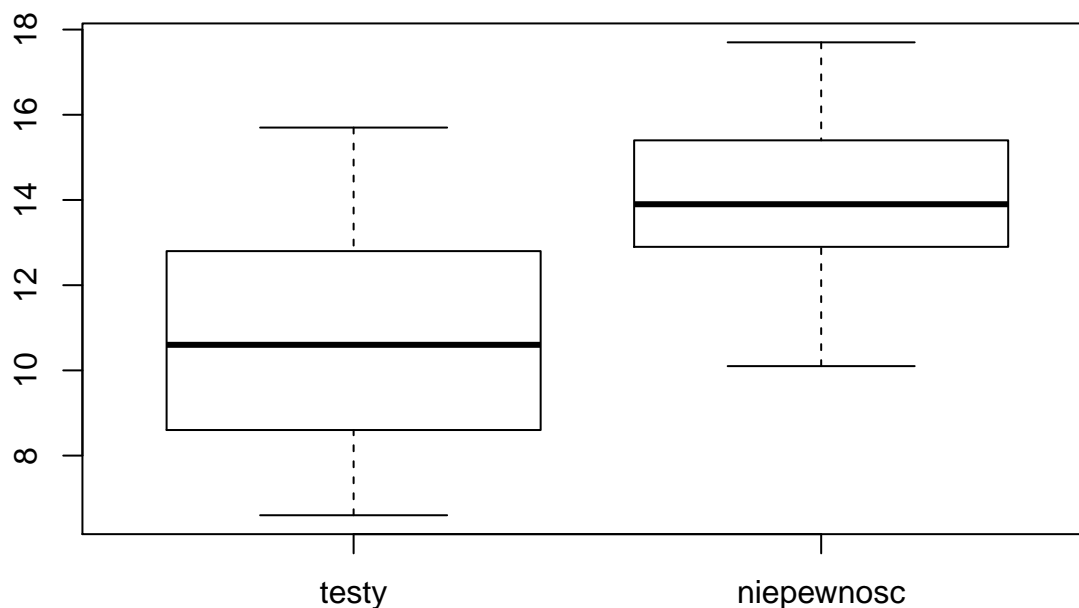
Aneta Przydróżna

Praca nad zbiorem danych, opisującym relacje między prawdopodobieństwem przyjęcia na studia a wynikami z testów rachunkowych i poziomu niepewności.

Dostanie się na studia oznaczało sukces czyli 1, natomiast porażka 0. Ze względu na binarne wartości zmiennej objaśniającej, użyję do analizy uogólnionych modeli liniowych, z kilkoma funkcjami linkującymi. Ale na początek graficzny zarys danych.



Na pierwszym wykresie, gdzie sukces zależy wyłącznie od wyników testów rachunkowych, widać, że osoby, które uzyskały wynik większy, bądź równy 12 zawsze zostawały przyjmowane, natomiast zależność nie zachodzi w drugą stronę, ponieważ sporo osób z wynikami w przedziale 8-12 również się dostała. Drugi wykres pokazuje przeciwną zależność: im wyższa niepewność tym mniejsza szansa przyjęcia. Jest on też bardziej klarowny, ponieważ wszystkie osoby, dla których niepewność wynosiła 10-13 dostały się, żadna za to nie dostała się na przedziale 16-18. Boxploty poniżej również wyglądają inaczej. Ten dla testów rachunkowych ma medianę niżej i jest znacznie szerszy niż boxplot dla niepewności. Stosunek szansy przyjęcia na studia do odrzucenia wynosi 1.38.



- Logit (regresja logistyczna)

$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$, gdzie $0 \leq p \leq 1$

	P-wartości	estymatory
intercept	0.0362	14.2386
numeracy	0.0200	0.5774
anxiety	0.0040	-1.3841

Dla hipotezy zerowej $\beta_1 = 0$, p wartość jest mniejsza niż 0.05, zatem ją odrzucamy i przyjmujemy, że parametr testów rachunkowych jest istotny w analizie. Dla $\beta_2 = 0$, p wartość jest znacznie mniejsza niż 0.05, więc można uznać, że parametr niepewności ma duże znaczenie w analizie. Kryterium AIC przyjmuje wartość 34.3.

- Probit

	P-wartości	estymatory
intercept	0.0247	8.2573
numeracy	0.0137	0.3371
anxiety	0.0014	-0.8039

P wartości dla hipotezy zerowej w tym modelu są odpowiednio do parametrów niższe niż w powyższym modelu, zatem również są istotne i przyjęta zostaje alternatywa. Kryterium AIC pokazuje 33.9 więc nieznacznie mniej niż powyżej.

- Cauchit

	P-wartości	estymatory
intercept	0.1350	18.3830
numeracy	0.1224	0.7323
anxiety	0.0735	-1.7741

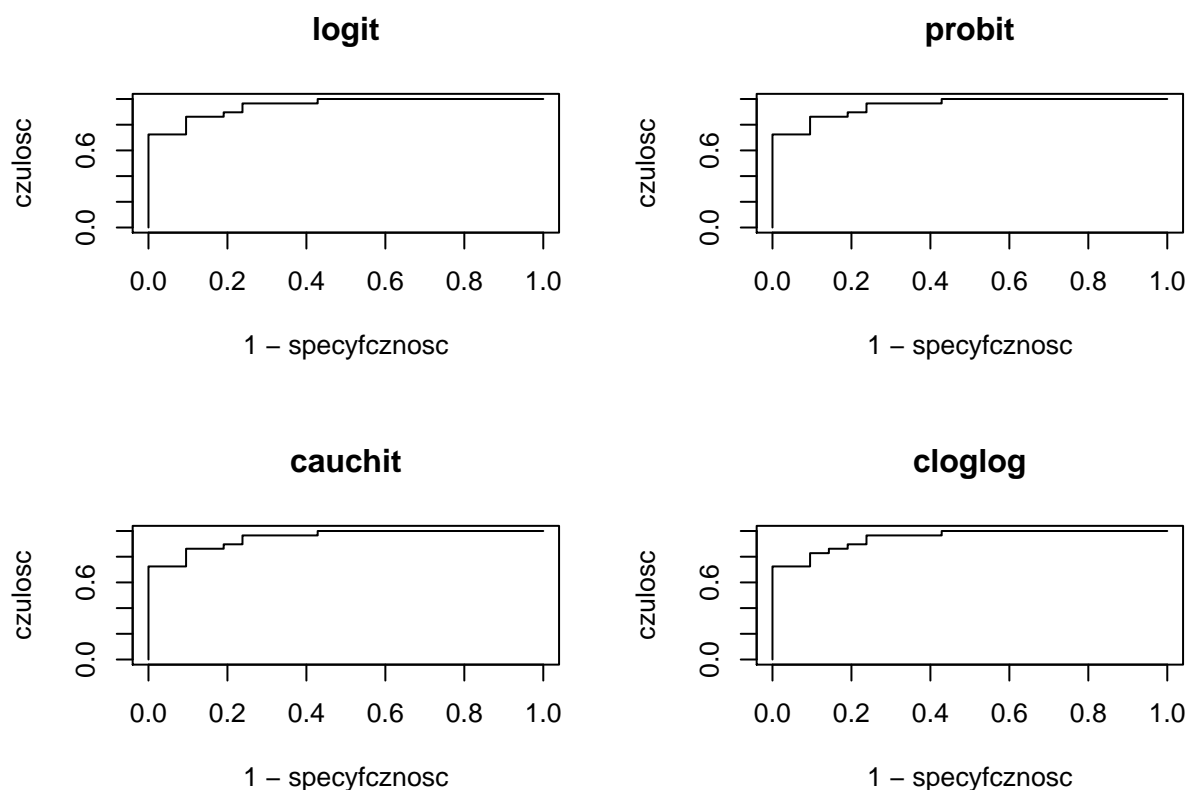
Dla hipotezy zerowej $\beta_1 = 0$, p wartość jest znacznie większa niż 0.05, zatem nie ma podstaw do odrzucenia hipotezy. Dla $\beta_2 = 0$, p wartość jest również większa niż poziom istotności, więc nie odrzucamy hipotezy. Kryterium AIC przyjmuje wartość 37.1.

- cloglog

	P-wartości	estymatory
intercept	0.0530	9.0006
numeracy	0.0082	0.4024
anxiety	0.0047	-0.9390

P wartości dla hipotezy zerowej w tym modelu są znacznie mniejsze niż 0.05 , zatem parametry są istotne i przyjęta zostaje alternatywa. Kryterium AIC pokazuje 34.

Jedynie model z funkcją Cauchit nie wykazał silnego związku parametrów z przyjęciem na studia, a kryterium AIC wskazało największą liczbę, więc można uznać, że model ten jest najgorszy. Model z Probit wykazał najmniejszą liczbę dla AIC, jest najkorzystniej dopasowany.



Jak widać powyżej, pierwsze 3 wykresy ROC są identyczne, natomiast czwarty różni się nieznacznie od pozostałych. Co pokazuje, że wszystkie funkcje linkujące dają podobne efekty.

Predykcja dla określonych zmiennych objaśniających

Przypisując wynikom testów rachunkowych liczbę 10, a poziomowi niepewności liczbę 13 otrzymałam następujące prawdopodobieństwa przyjęcia dla każdej funkcji. Wszystkie są podobne, z różnicami zaczynającymi się na poziomie dwóch miejsc po przecinku. Największe prawdopodobieństwo dostania się na studia, przy określonych zmiennych, wykazała funkcja cloglog a najmniejsza funkcja probit.

logit	probit	cauchit	cloglog
0.88279	0.88060	0.88485	0.89630