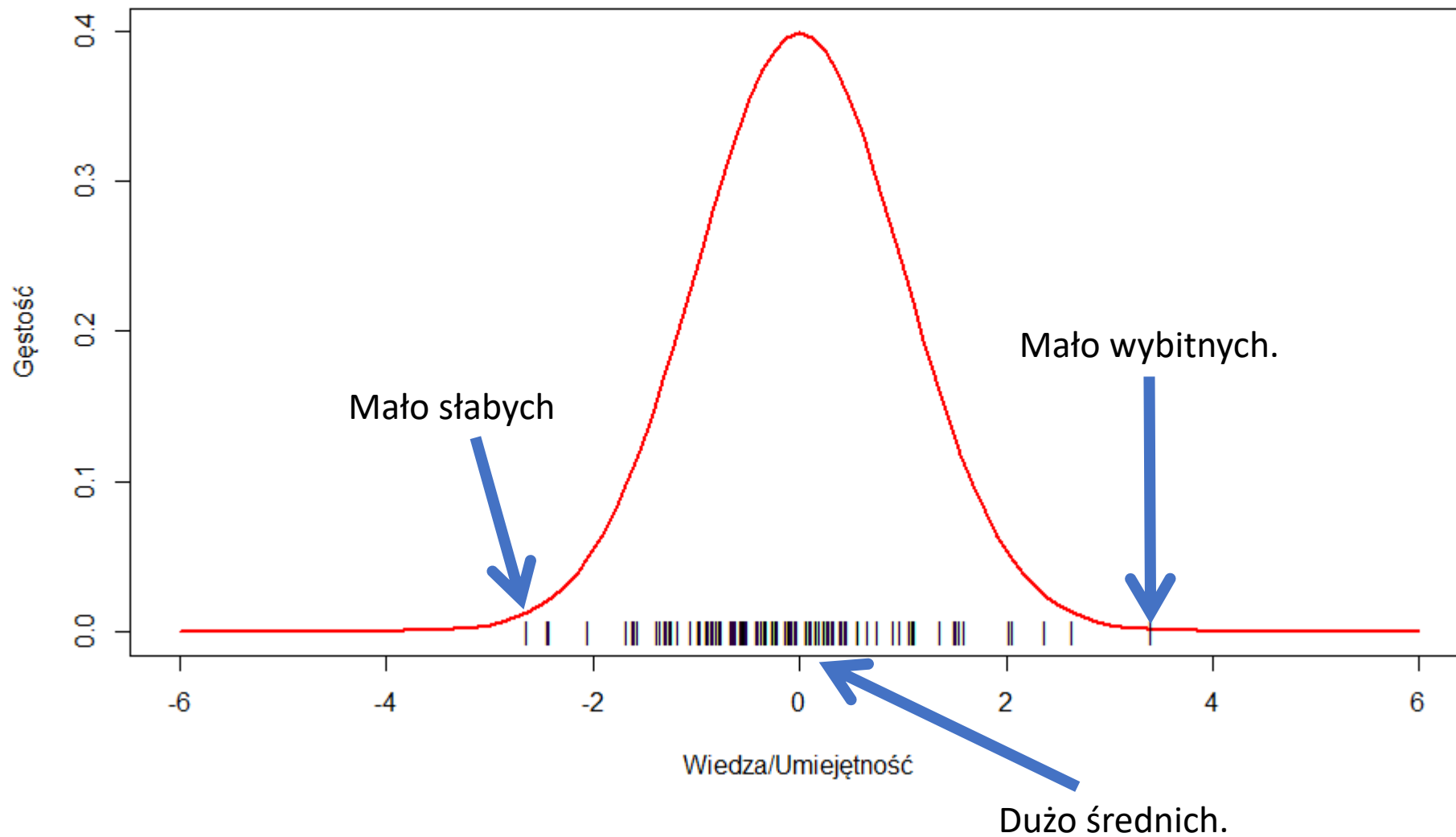


# Nie męczymy tak uczniów!

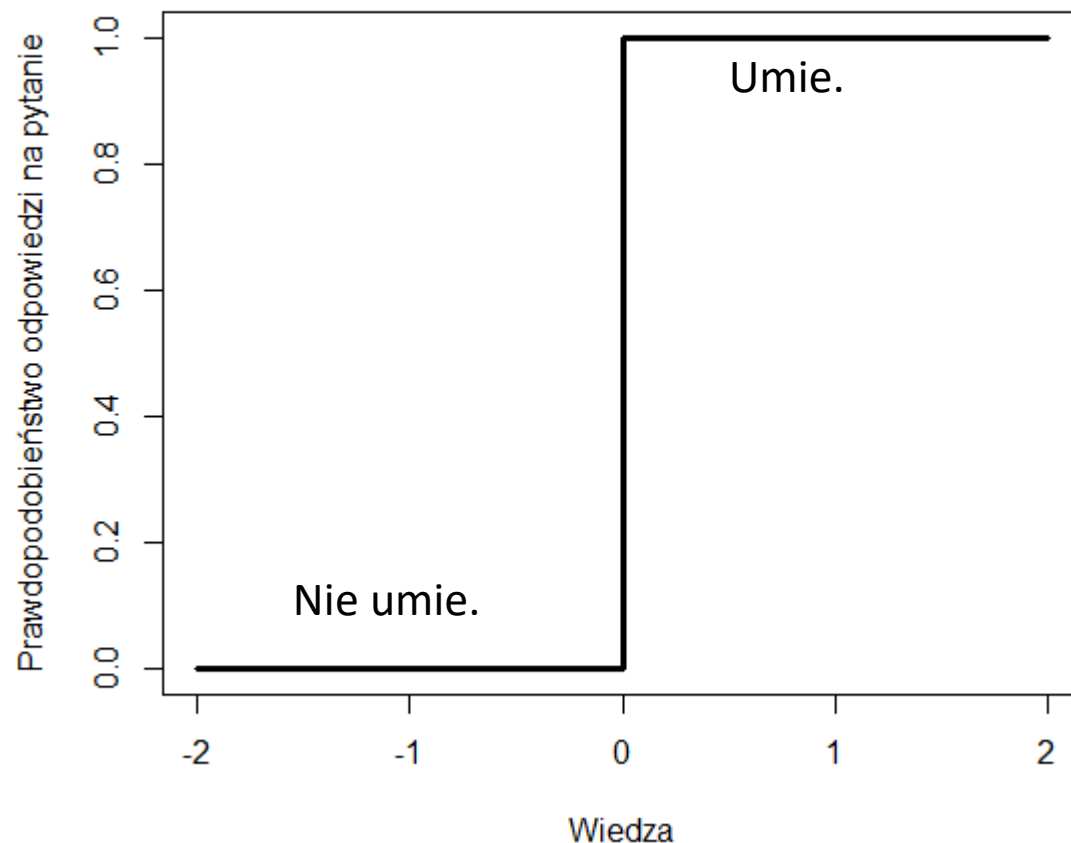
## Praktyczne zastosowanie modelu IRT



## Rozkład umiejętności wśród uczniów:



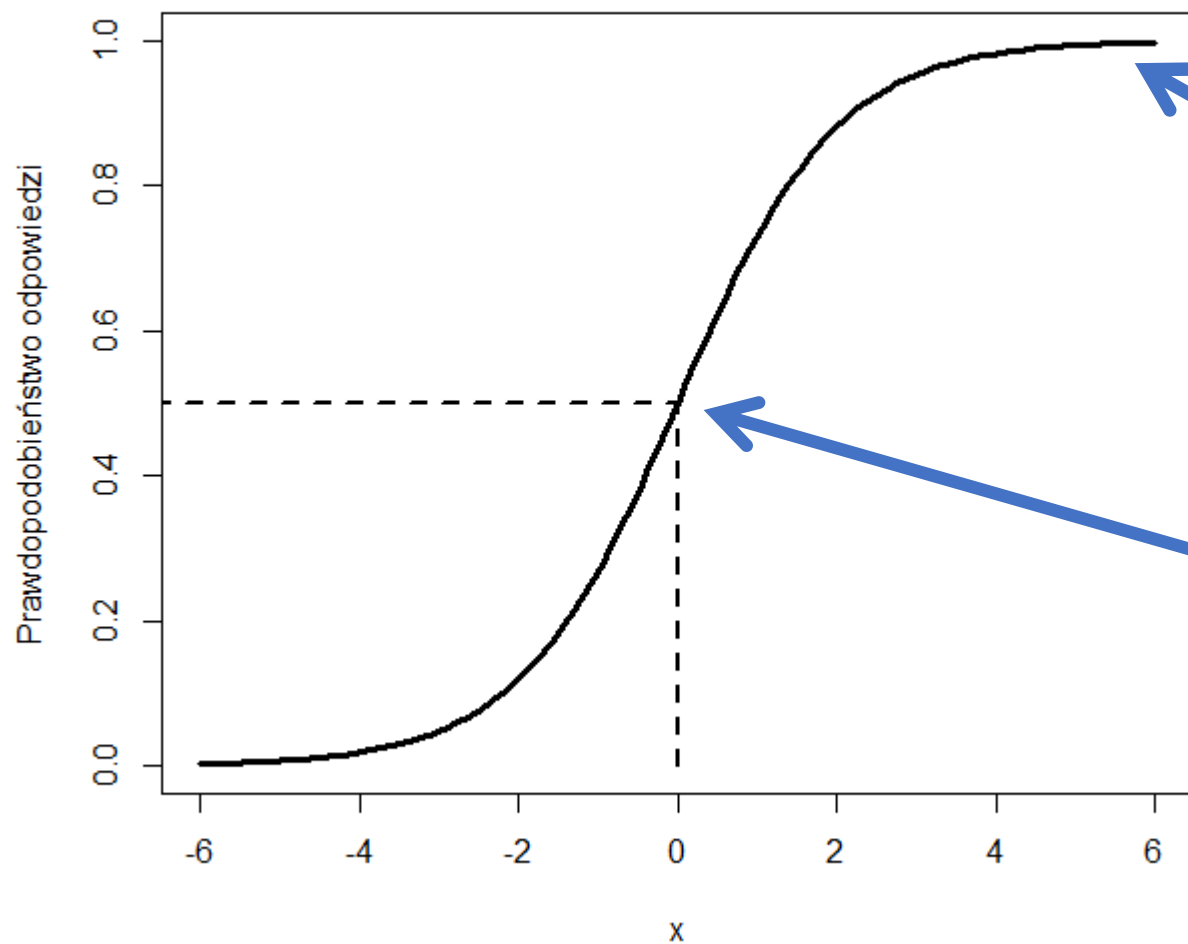
Testując wiedzę można  
przyjąć model postaci:



← Ale oczywiście  
tak nie jest!

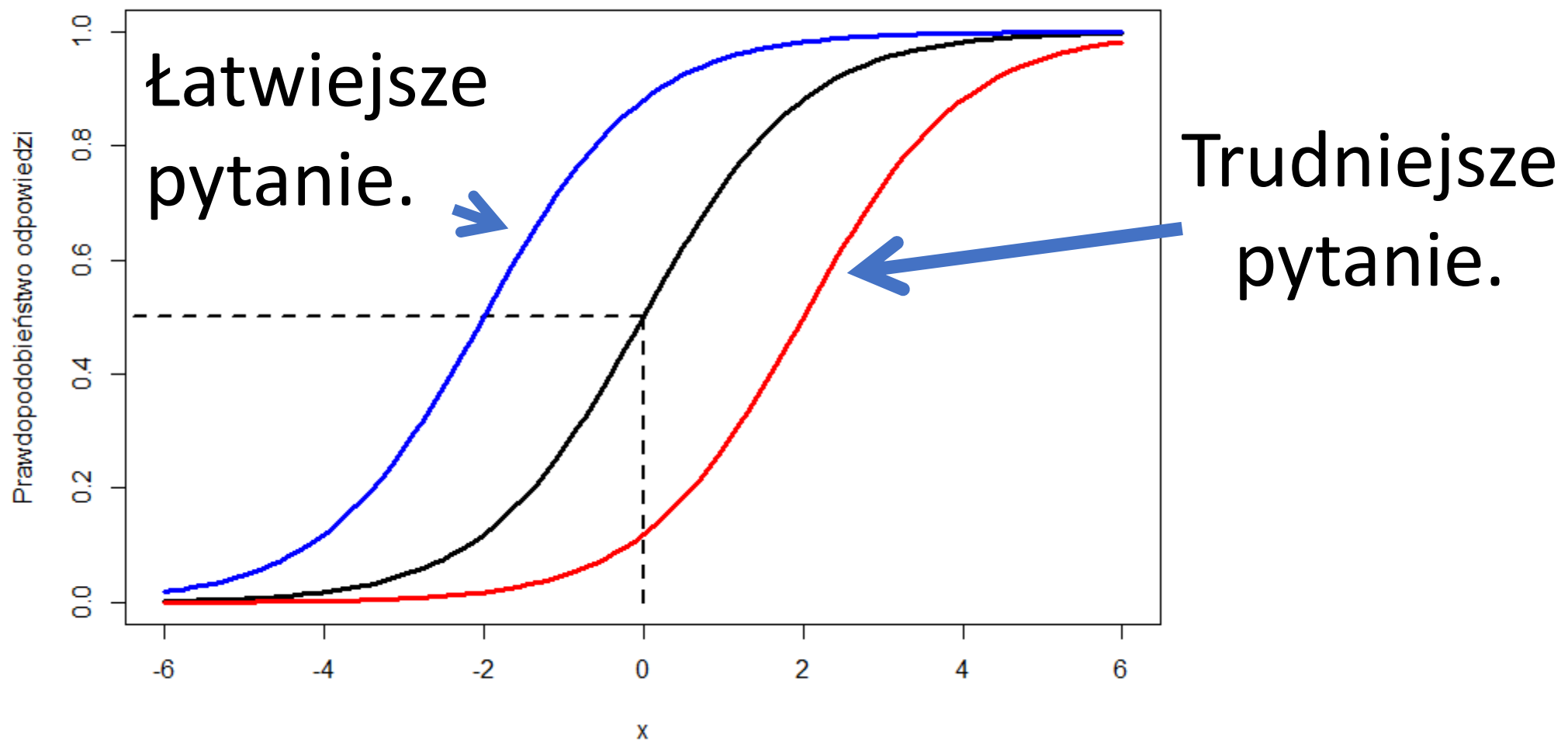
Uczeń może źle odpowiedzieć, gdyż:

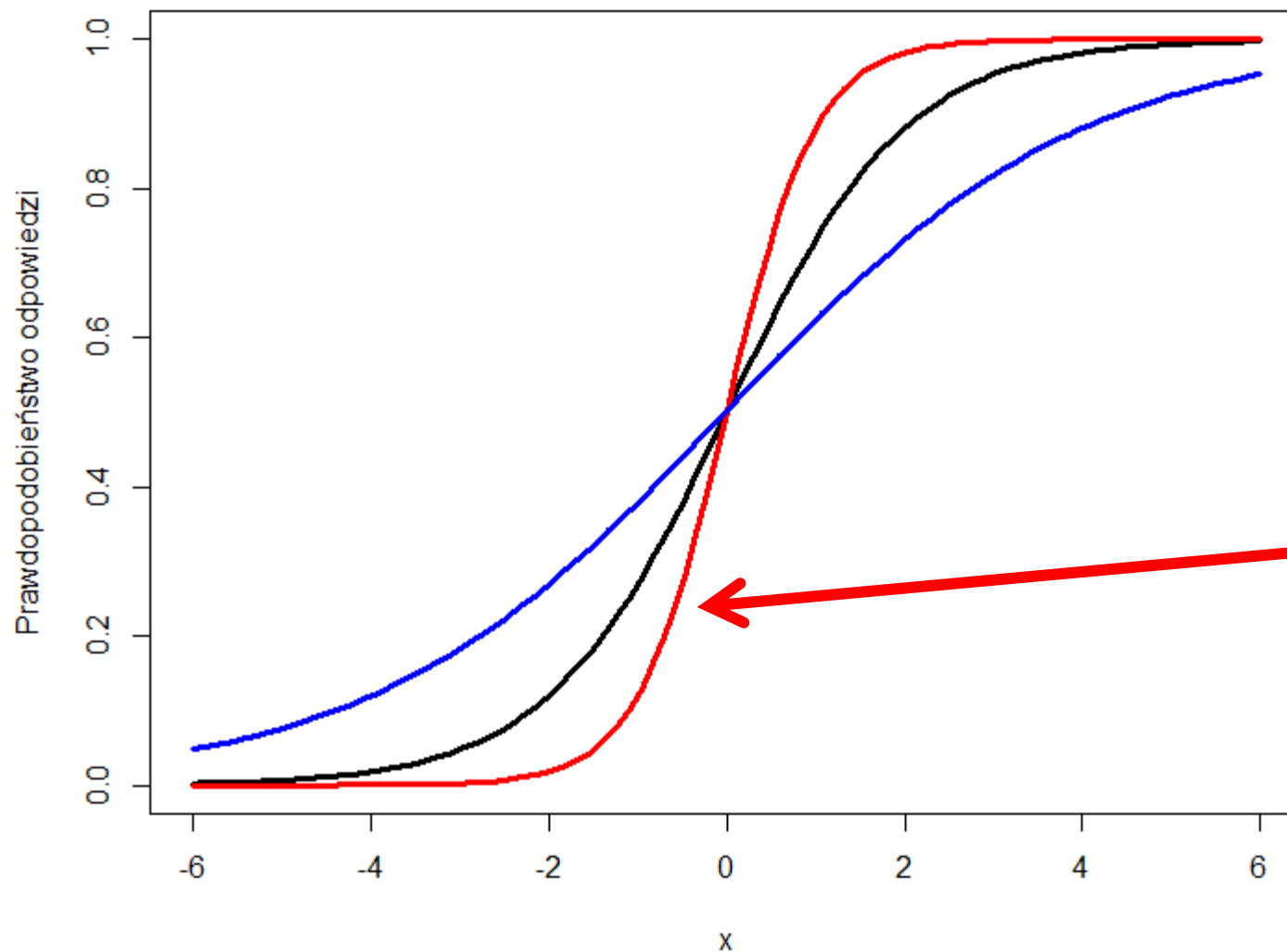
- ma gorszy dzień,
- coś go rozproszyło,
- za oknem świeci słońce,
- itp.



Zdolny uczeń  
odpowie **prawie**  
na pewno.

Średni uczeń ma jakieś  
prawdopodobieństwo  
odpowiedzi.

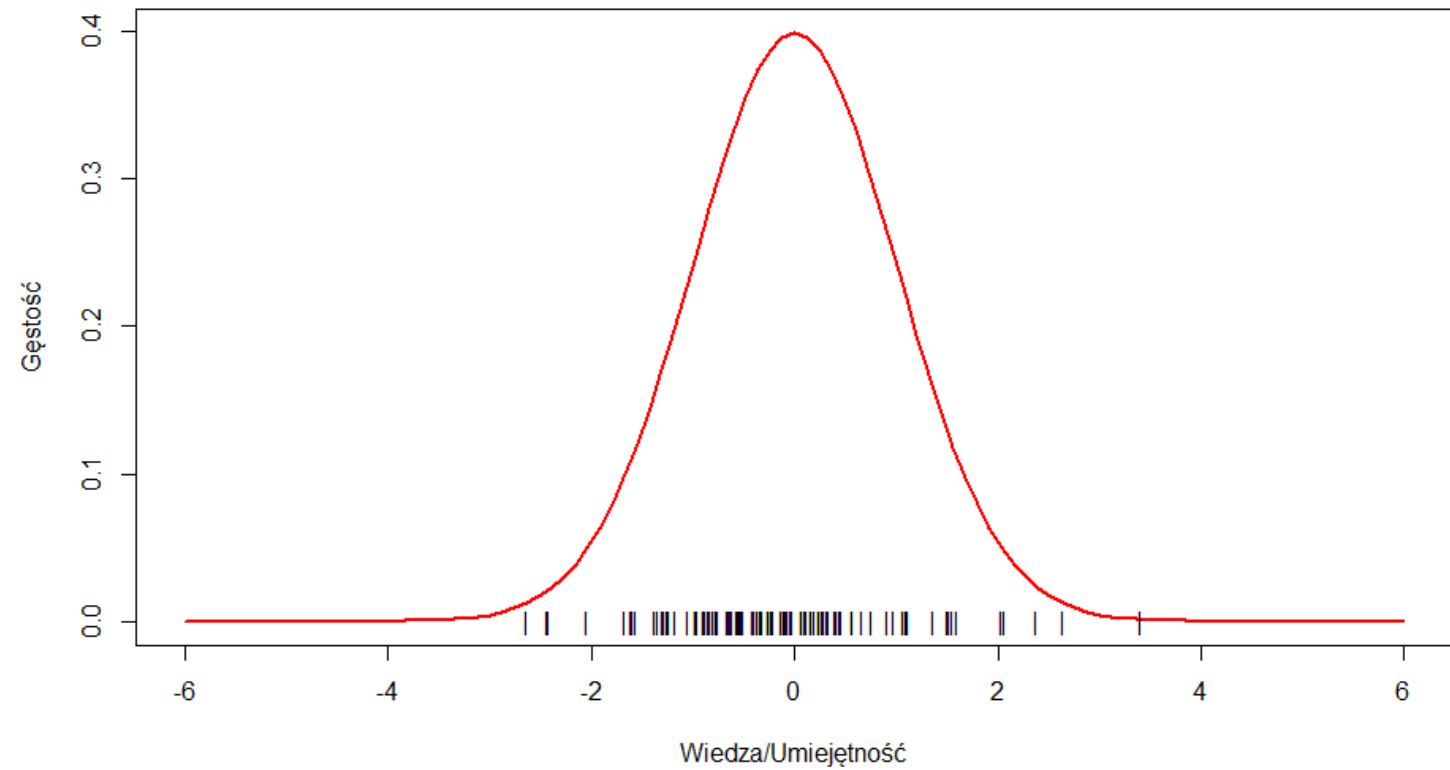




Bardziej stroma  
krzywa – mała  
zmiana zdolności  
powoduje dużą  
zmianę szans na  
poprawną  
odповідź


Dla przypomnienia:

## Rozkład zdolności:



Wzory:

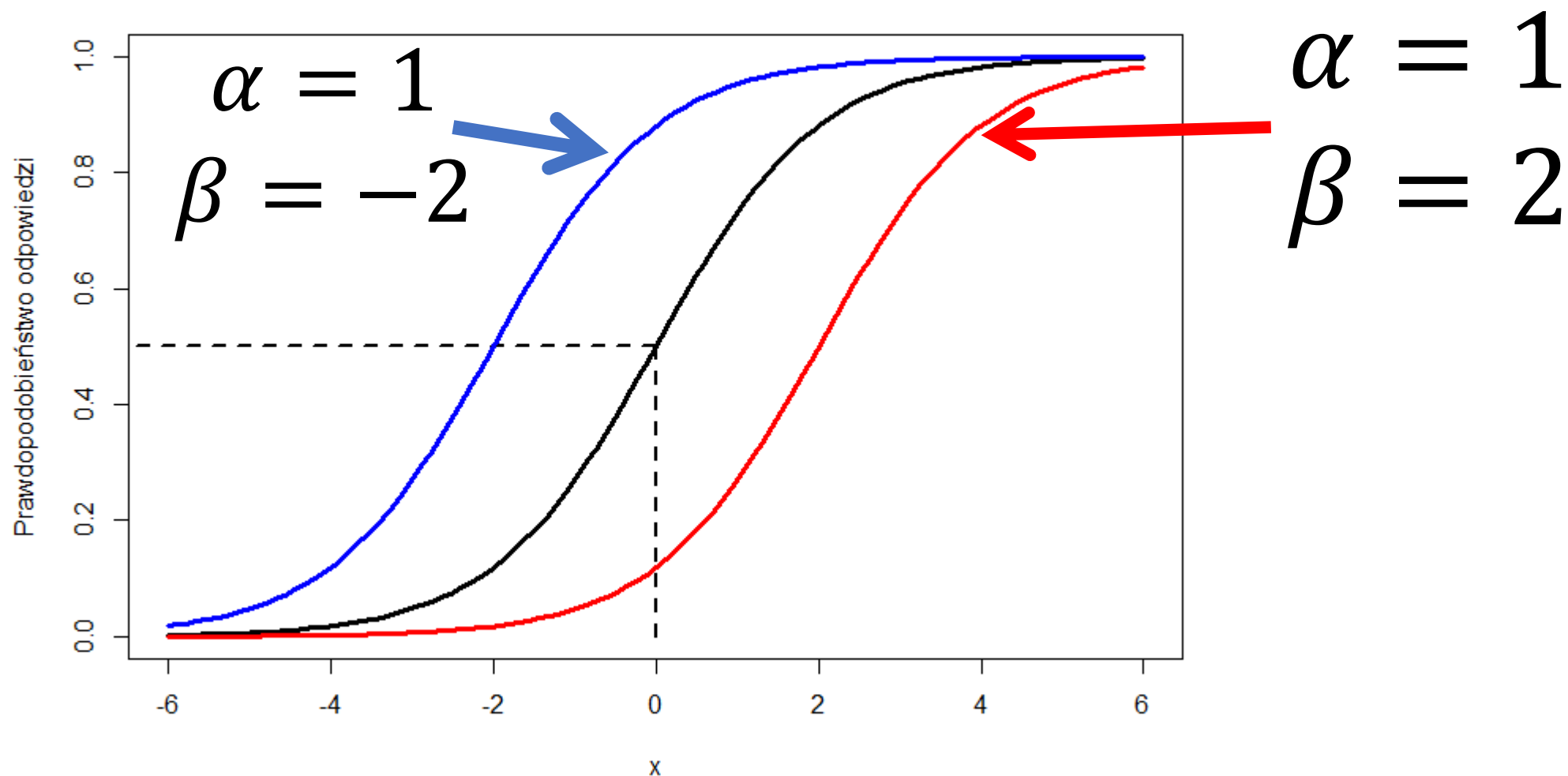
$$P(x) = e^{\alpha(x - \beta)} / (1 + e^{\alpha(x - \beta)})$$

Ciężki wzór... 

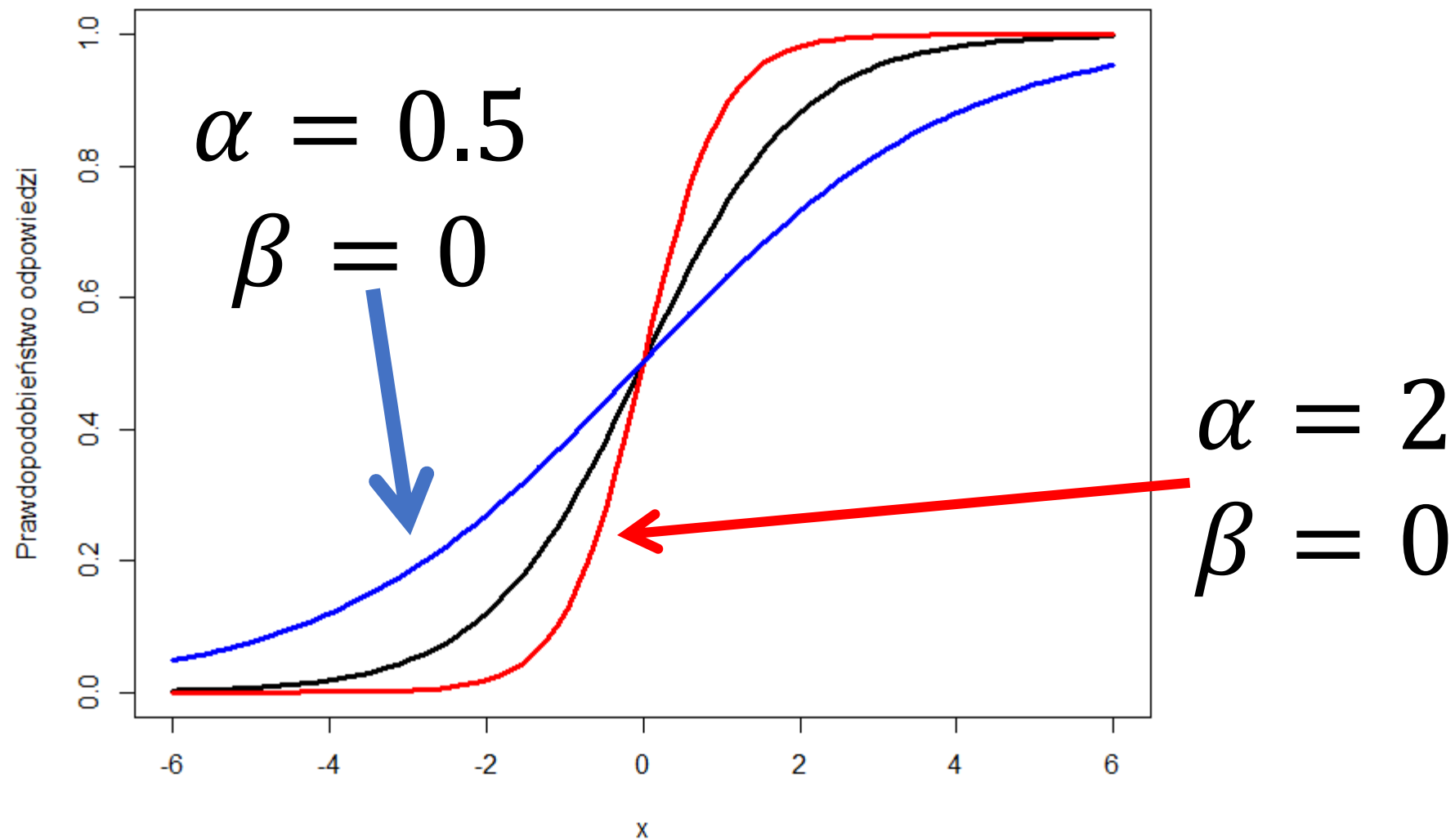
$\beta$  – określa jak trudne jest pytanie.

$\alpha$  – określa zdolność różnicowania pytania.





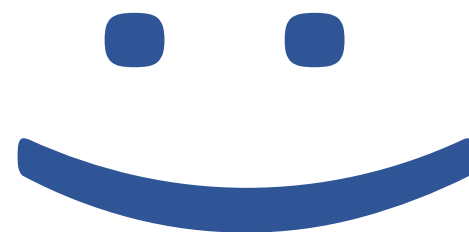
$$P(x) = e^{\alpha(x - \beta)} / (1 + e^{\alpha(x - \beta)})$$



# Model IRT...

Nareszcie!

...w dużej mierze został już  
przedstawiony!

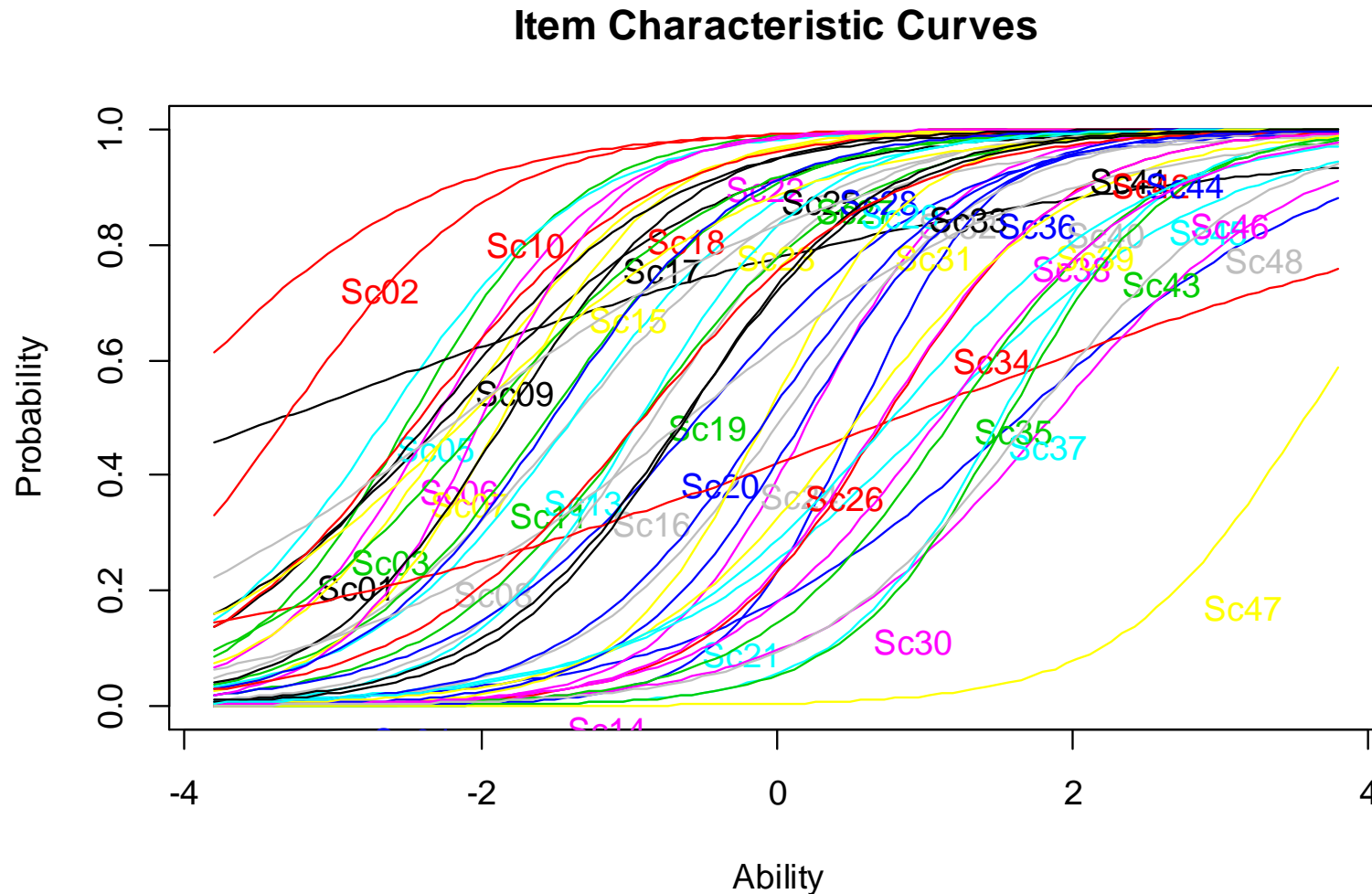


Nie wchodząc w szczegóły techniczne  
– dopasowanie modelu IRT polega  
na znalezieniu wartości  $\alpha$  i  $\beta$  dla  
każdego pytania.

Teorię można poczytać w domu...

# Oszacowany model IRT:

Krzywa  
charakterystyczna  
dla każdego  
pytania.



# Problem:

## Dziesiątki tysięcy uczniów do sprawdzenia

Oszczędność czasu się liczy!

# W czym IRT może pomóc?

By wyznaczyć zdolności ucznia, nie musi on odpowiadać na wszystkie pytania!



Co prawda im więcej pytań wchodzących w skład testu tym lepsza dokładność wyniku...

Można szacować zdolności po każdej odpowiedzi!



A gdyby tak przeprowadzać test adaptacyjnie?

Po każdym pytaniu system  
dobiera następne bazując  
na aktualnym  
oszacowaniu zdolności  
ucznia.

Poprawna  
odpowiedź



Trudniejsze  
pytanie.

Błędna  
odpowiedź



Łatwiejsze  
pytanie.

# Informacje danych:

- W każdym pytaniu można zdobyć 1 lub 0 punktów (dobrze/źle).
- 9297 uczniów ze szkoły zawodowej z okolic Johannesburga (RPA).
- 11 języków narodowych – jednak lekcje prowadzone są po angielsku.
- Test składa się z 48 pytań.

Trochę R:

Dane:

Pakiet:

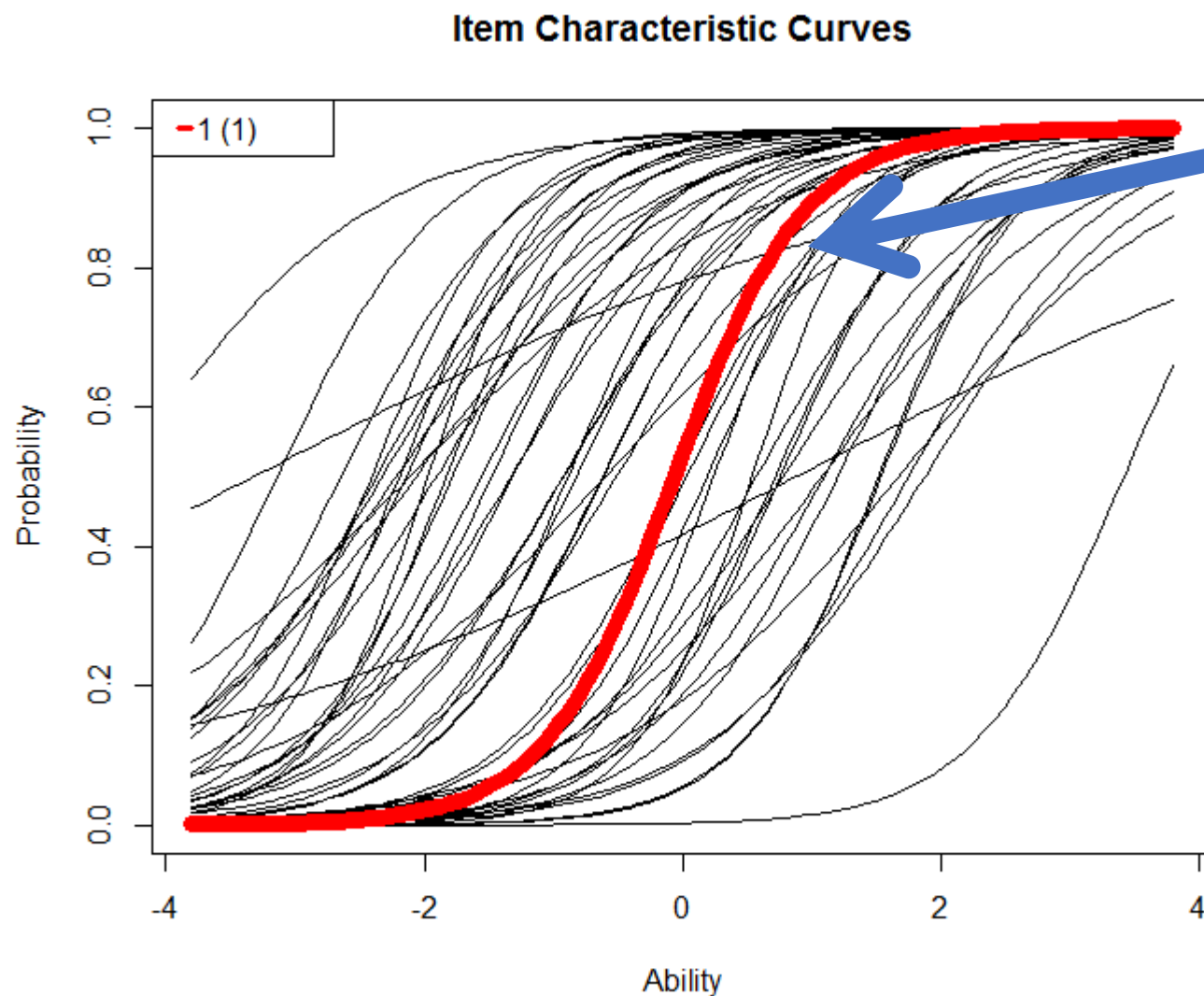
ltm

Kod:

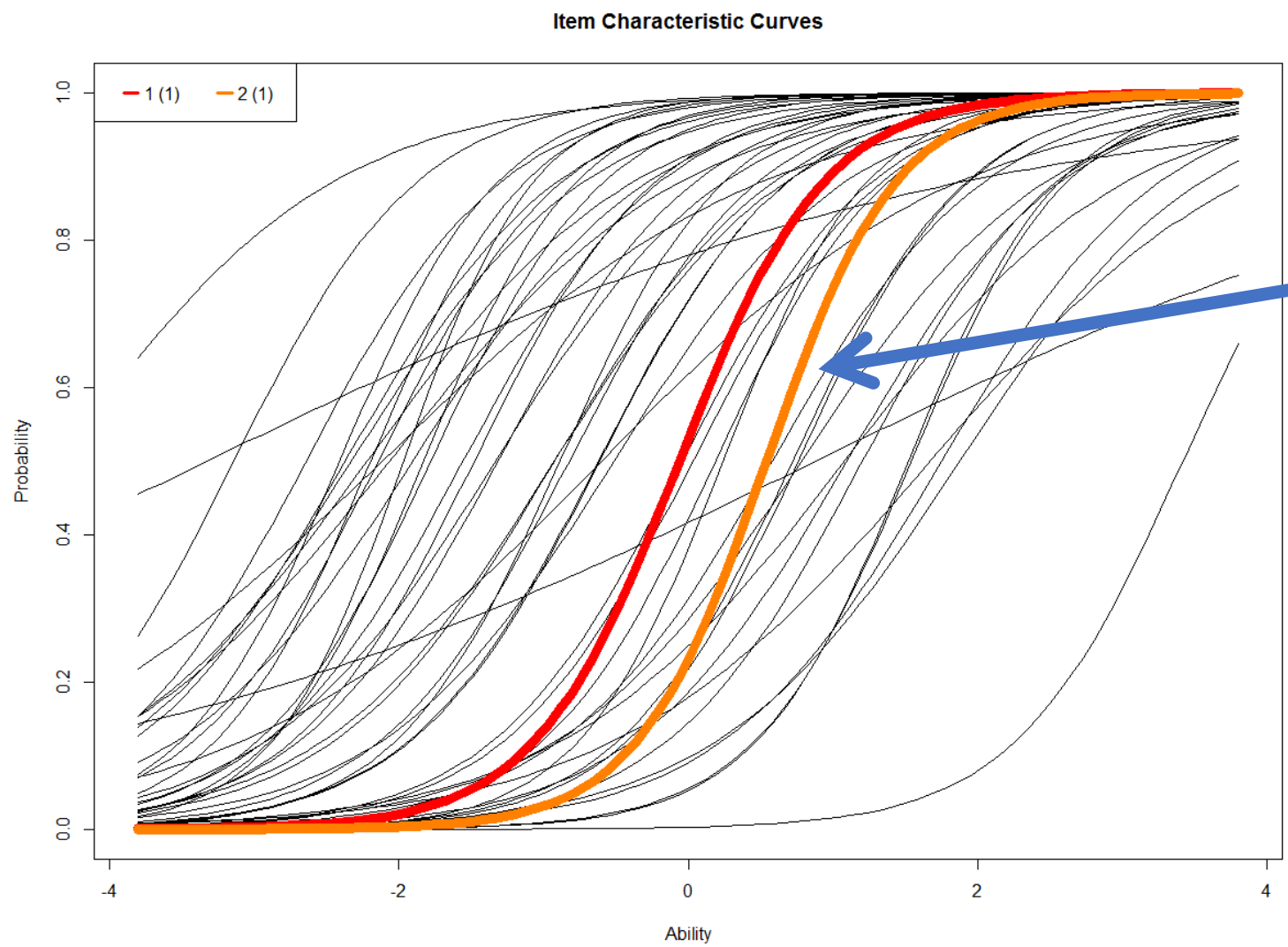
	Sc01	Sc02	Sc03	Sc04	Sc05	Sc06	Sc07	Sc08	Sc09	Sc10	Sc11	Sc12	Sc13	Sc14	Sc15	Sc16	Sc17	Sc18	Sc19	Sc20	Sc21	Sc22	Sc23	Sc24
1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0
2	0	1	1	0	1	1	1	0	1	1	1	0	1	0	1	0	1	1	1	1	0	1	1	0
3	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0
4	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1	1	0
5	1	1	1	0	1	0	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0
6	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
8	1	1	1	0	1	1	1	0	0	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1
9	0	1	0	0	1	1	1	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	1	0
10	0	1	1	0	1	1	1	0	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	0
11	1	1	1	0	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	0	1	1	1	0
12	1	1	1	0	1	1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	0	1	1	0
13	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
14	0	1	1	0	1	1	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1	0	0

```
12  
13 model.fit <- ltm(train.data.scores ~ z1, IRT.param = TRUE)  
14 all.scores <- factor.scores(model.fit, test.data.scores)  
15
```

# Przykładowa sesja ucznia:

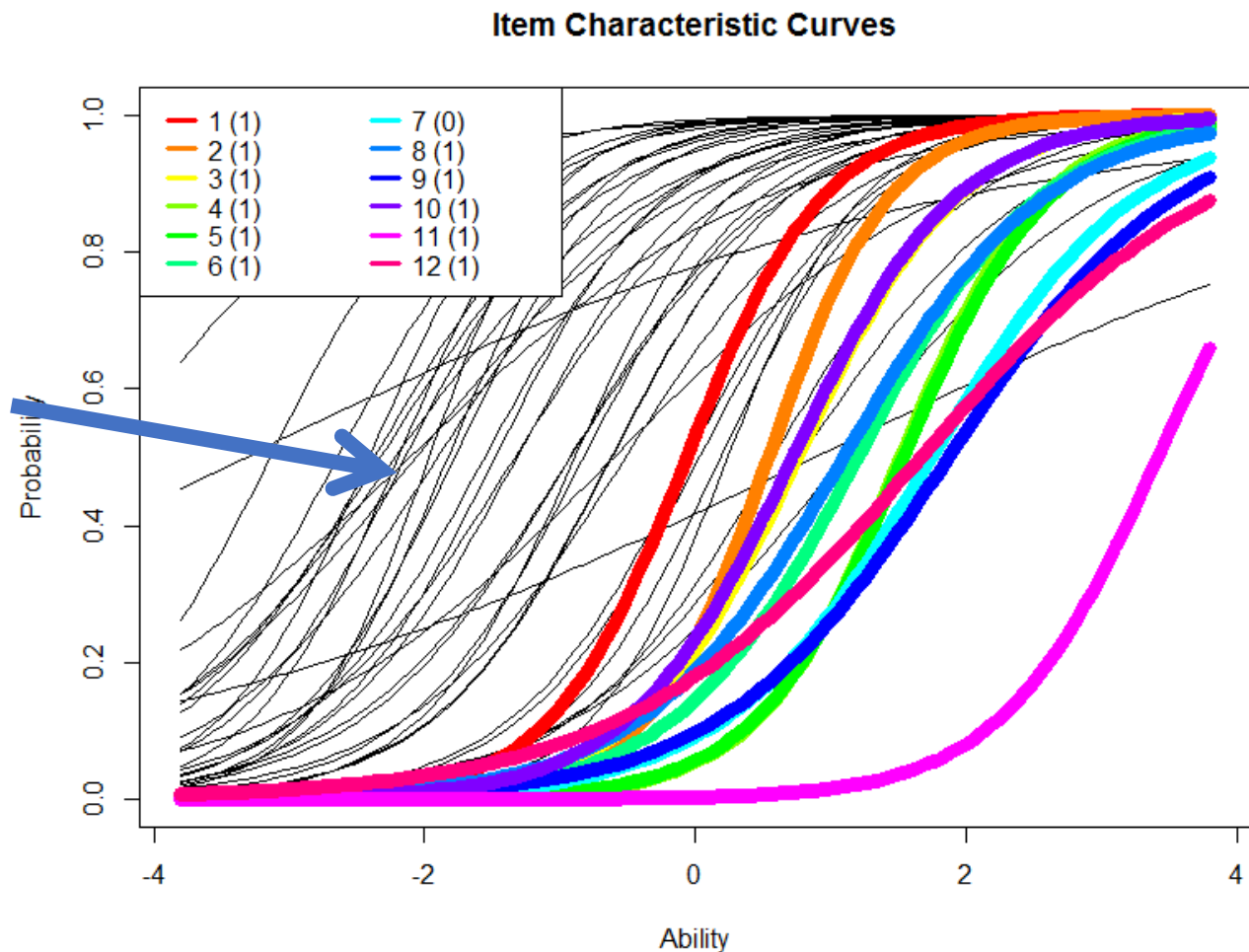


Pierwsze  
pytanie.

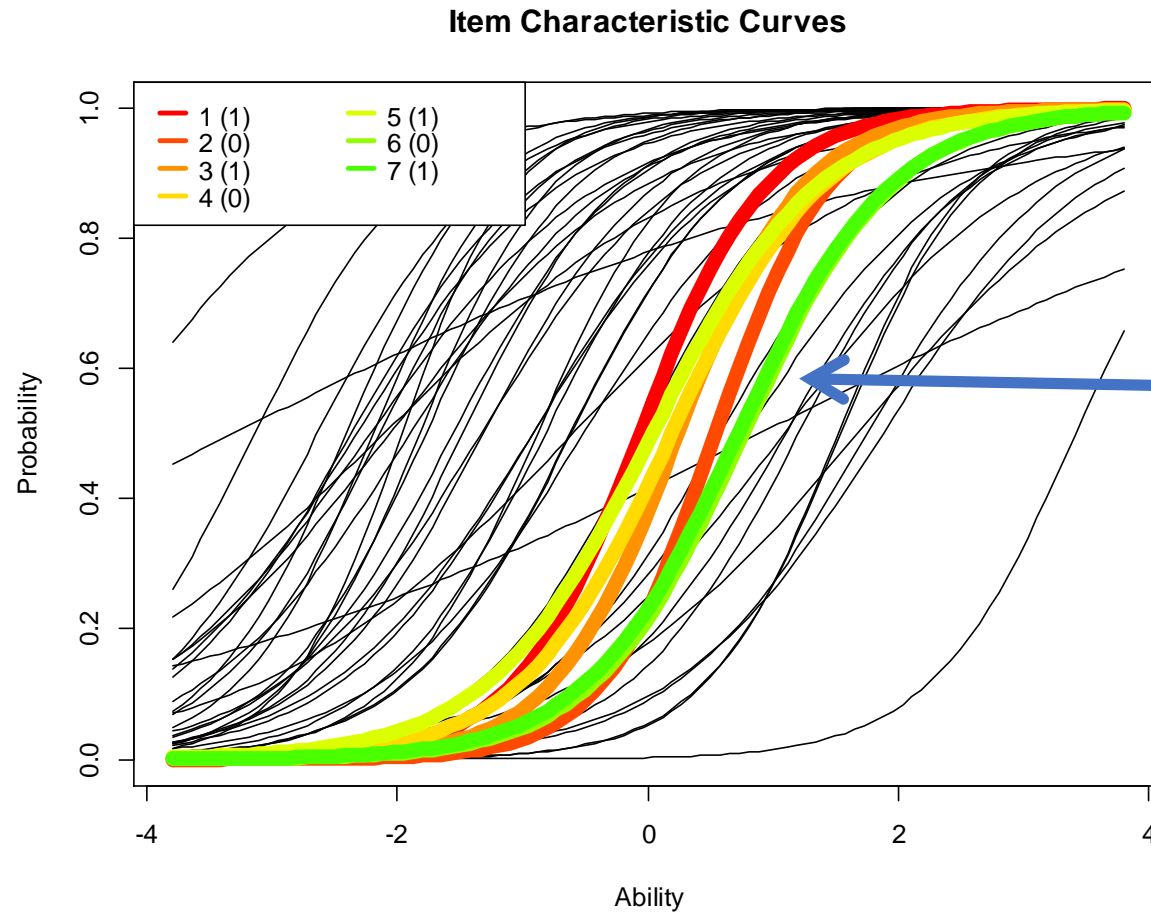


Drugie pytanie.

Cała sesja. Co ciekawe  
– uczeń ten nie dostał  
żadnego pytania bardzo  
łatwego!



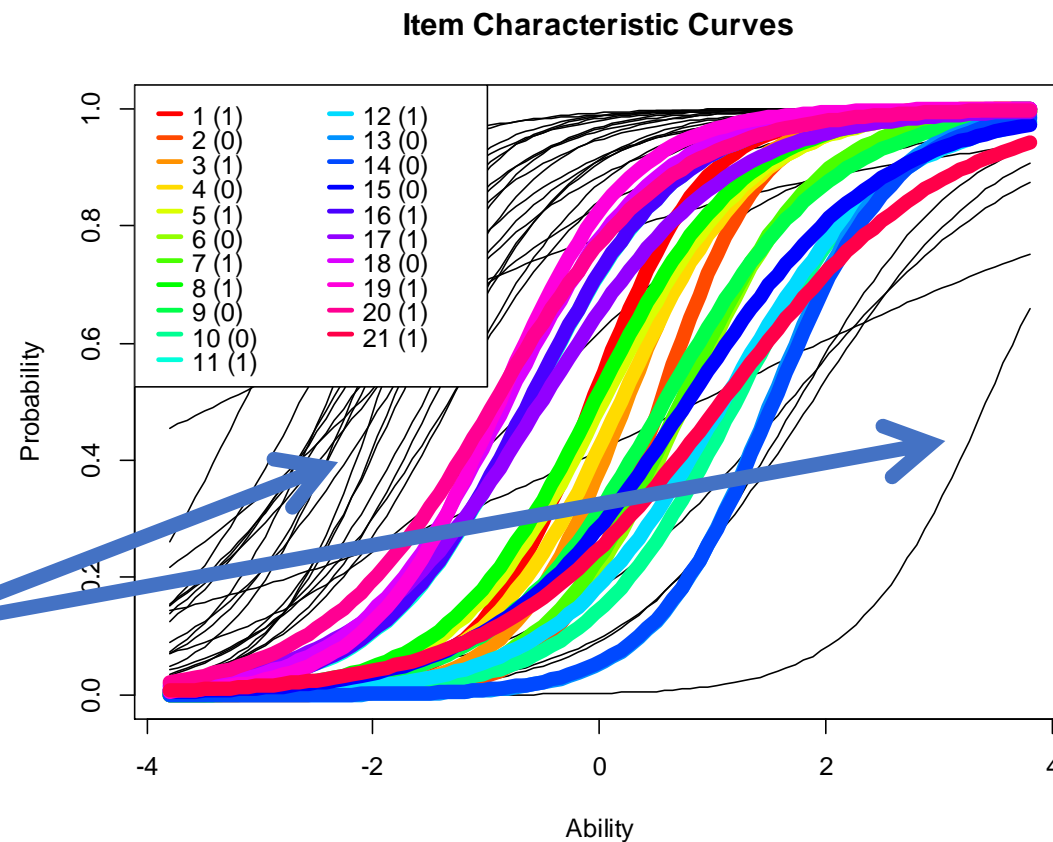
# Średni uczeń:



Początek sesji - pytania o  
zbliżonym poziomie  
trudności.

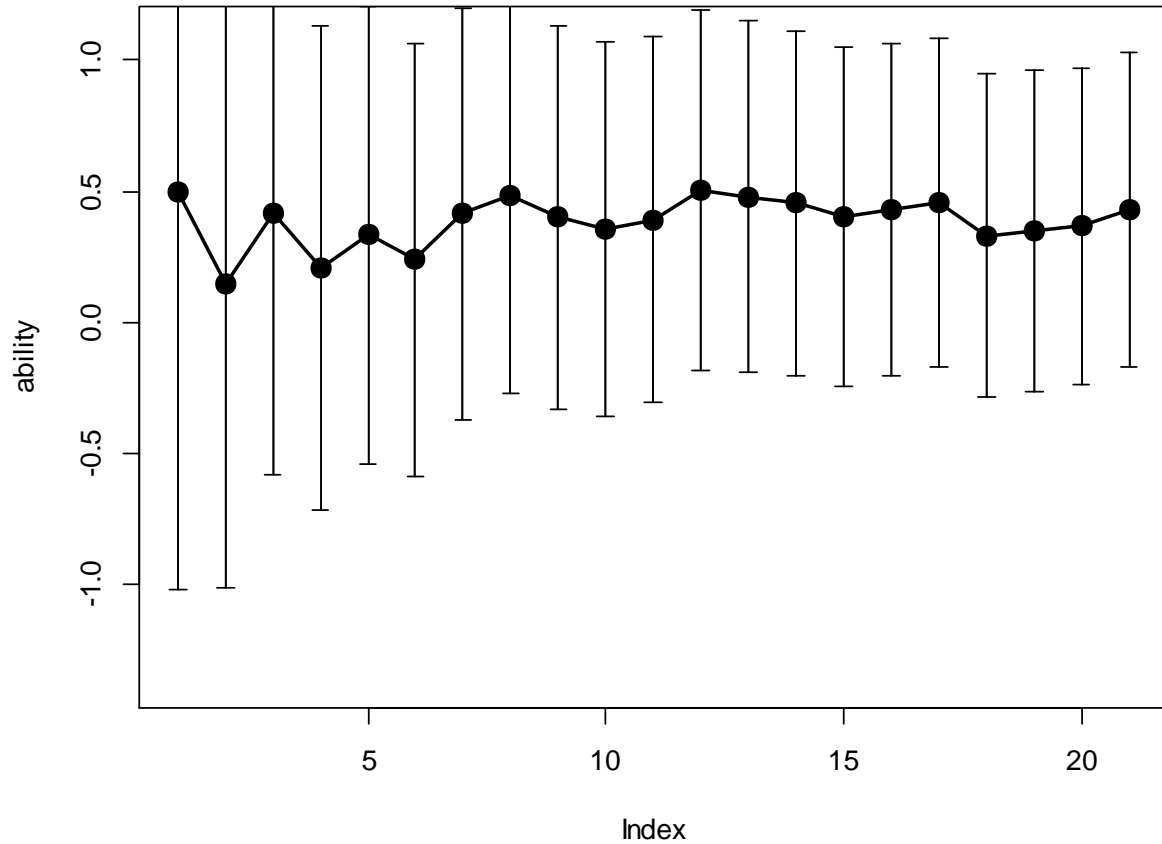
# Cała sesja:

**Najłatwiejsze i  
najtrudniejsze pytania w  
ogóle nie zostały  
zadane!**

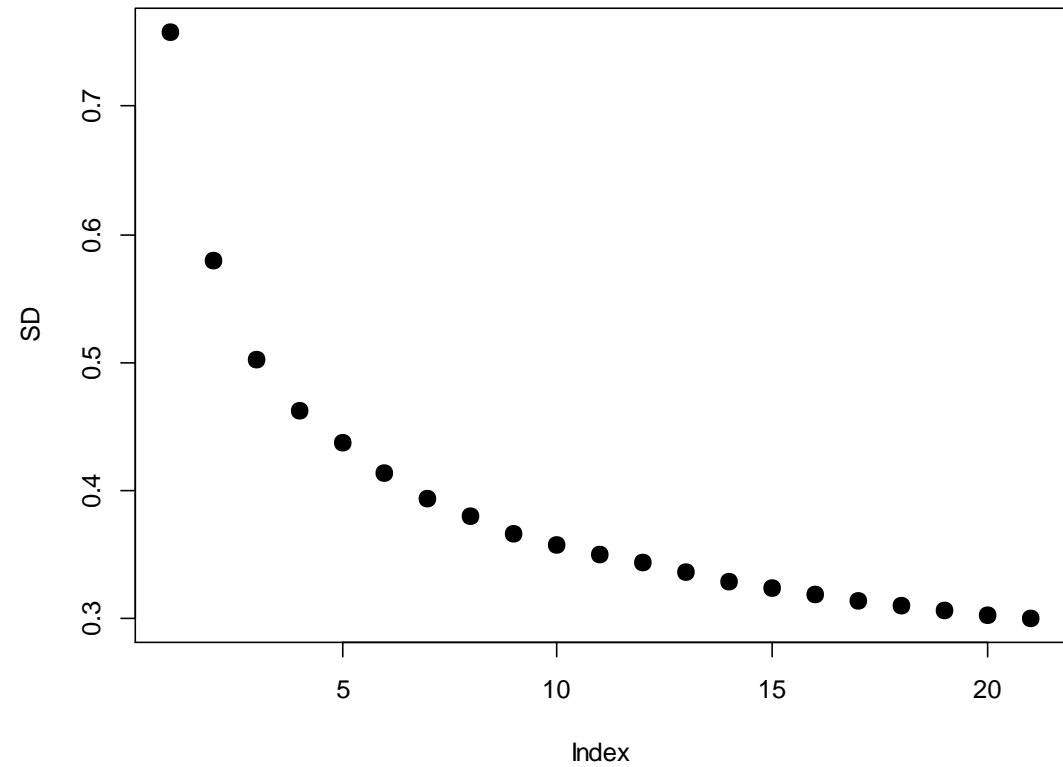




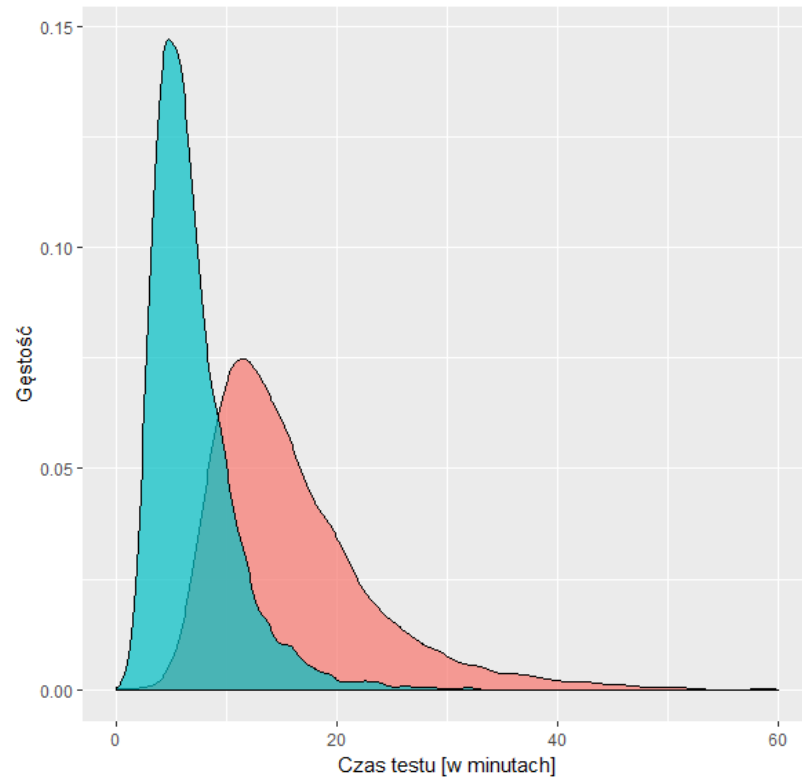
# Przebieg testu z innej perspektywy:



Z każdym kolejnym pytaniem  
SD wyniku spada

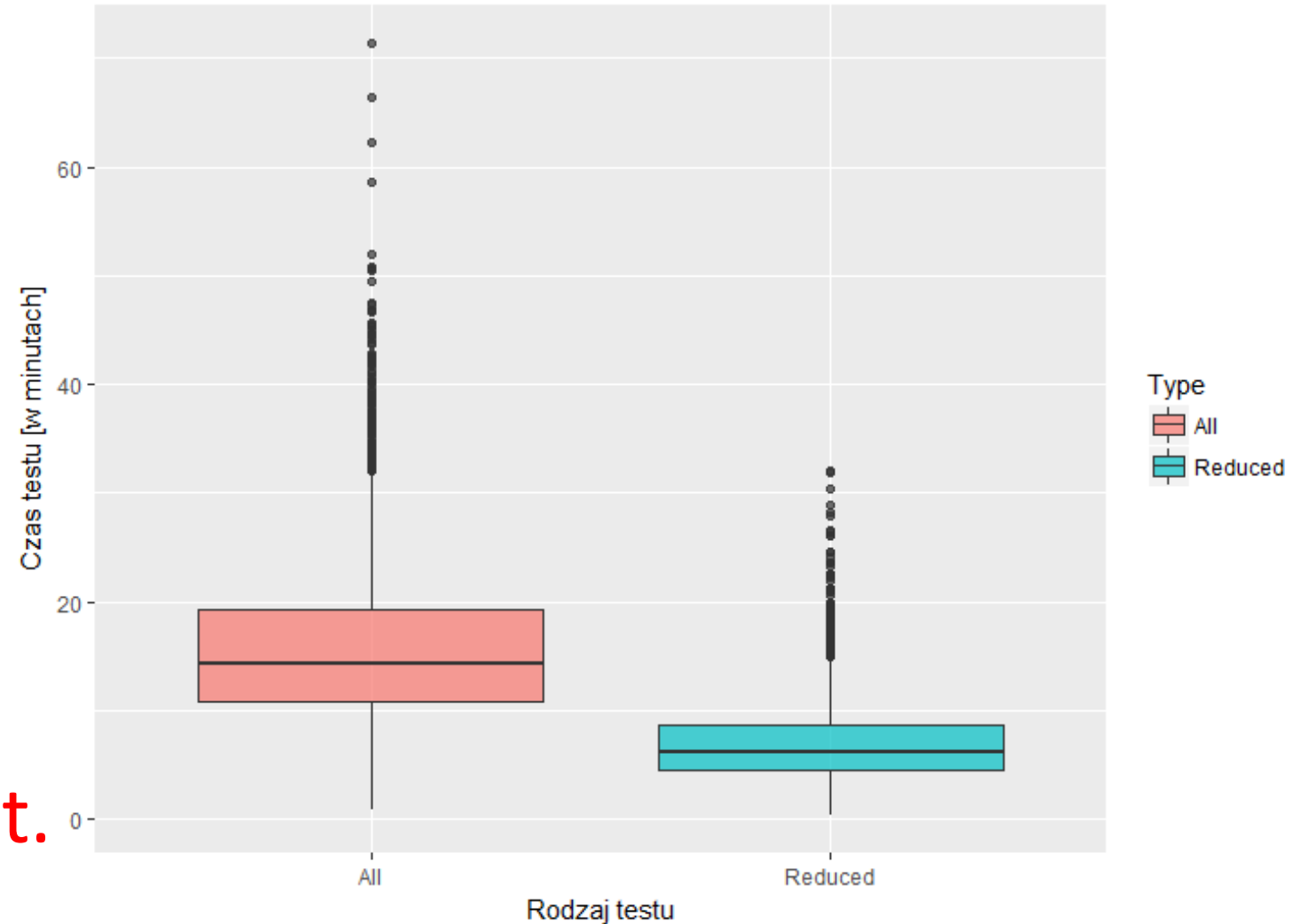


# Redukcja czasu trwania testu:



Średnia przed: 16 minut

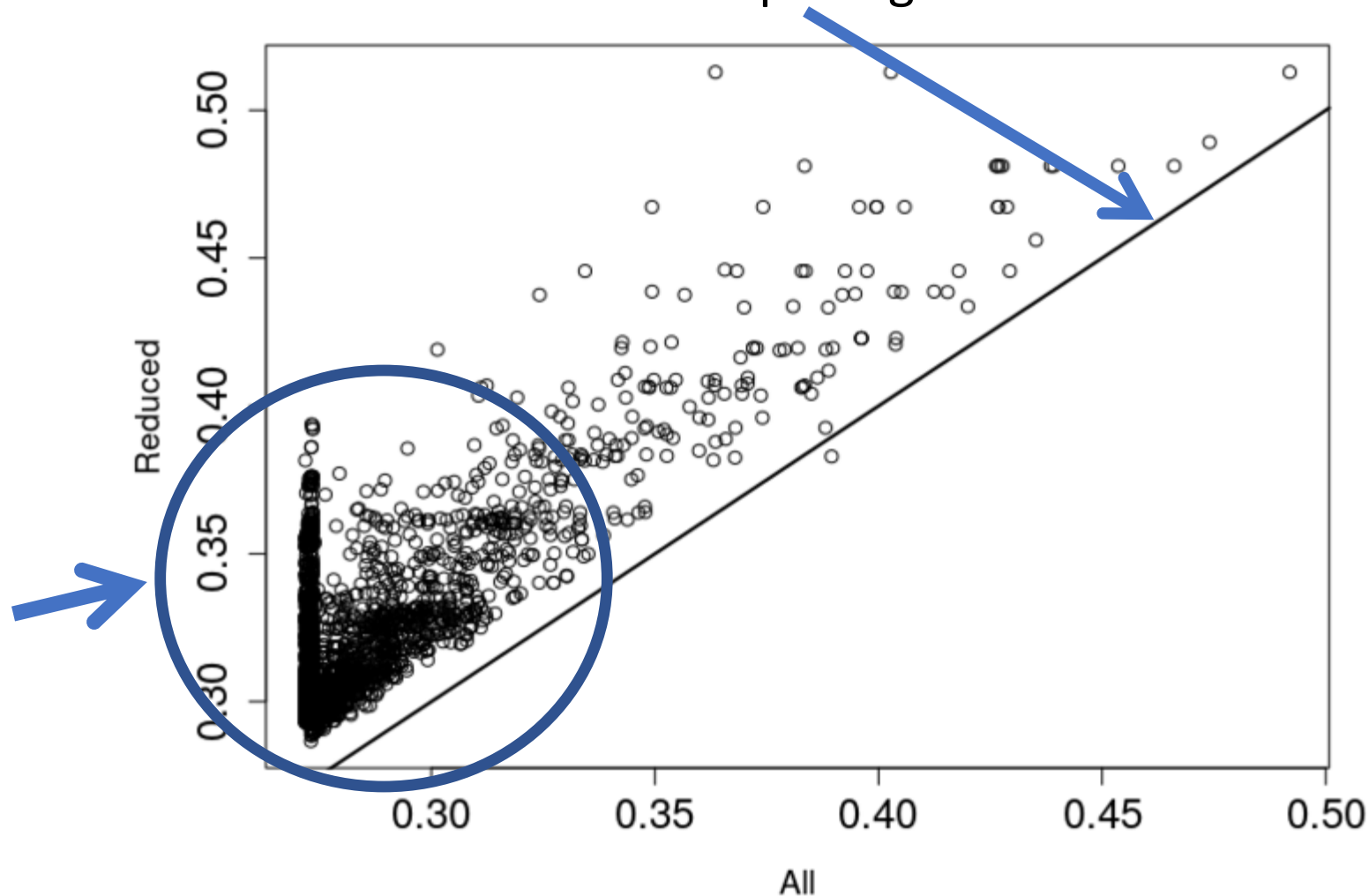
Średnia po: 7 minut.



# Koszt redukcji:

Na tej linii SD z modelu zredukowanego jest takie samo jak w przypadku modelu pełnego.

SD w modelu zredukowanym większe o kilkanaście %



To dopiero pierwsza iteracja modelu...

The biggest room in the  
world is the room for  
improvement.

# Dalsze kroki...

- Dokładna weryfikacja założeń modelu.
- Uwzględnienie w modelu cech charakterystycznych dla języków narodowych (typowe błędy – 'map' -> 'mip').
- Zmiany w metodologii wyznaczania końca testu.

# Dziękuję za uwagę!



# Czas na pytania!

W trakcie można też  
zanotować kontakt:



Kontakt e-mail:  
[zygmunt@zstat.pl](mailto:zygmunt@zstat.pl)