XXII KDE: Kraków 23.09.2016

Warsztaty
Pomiar edukacyjny i psychologiczny w programie R dla zaawansowanych

Karolina Świst i Tomasz Żółtak

IRT a KTT

IRT a KTT

- **Klasyczna Teoria Testu**
  - Rozwijana od przełomu XIX i XX w.
  - Punktem wyjścia pomiary fizykalne (**zmienne ciągłe**).
  - Centralny problem rzetelności (w kontekście *poprawki na ściągnięcie* stosowanej do korelacji).
  - Skoncentrowana na własnościach oszacowania mierzonej cechy.
  - W postaci analizy czynnikowej ważna kwestia analizy wielowymiarowej.
  - Jednocześnie długo pomijany problem nieadekwatności założeń klasycznej analizy czynnikowej (dla zmiennych ciągłych) do modelowania danych zawierających zmienne porządkowe (w szczególności wyniki rozwiązania zadań ocenianych binarnie).

IRT a KTT

- **Teoria Odpowiedzi na Zadanie Testowe (IRT)**
  - Skoncentrowana na modelowaniu zależności pomiędzy mierzoną cechą a wynikami rozwiązania poszczególnych zadań (**zmienne porządkowe**).
  - U zarania związki z badaniami toksykologicznymi (wpływ wielkości dawki na prawdopodobieństwo wystąpienia zdarzenia).
  - Przez bardzo długi czas (niemal do końca XX w.) w praktyce zajmująca się tylko testami jednowymiarowymi.
    - Skutek technicznych problemów z estymacją.
  - Formalne relacje pomiędzy podstawowymi pojęciami KTT i IRT opracowane wyczerpująco pod koniec lat 60. XX w. (Lord i Novick, 1968).
  - Formalne relacje pomiędzy analizą czynnikową estymowaną na podstawie macierzy korelacji polichorycznych a modelami IRT opracowane w II poł. lat 80. XX w. (Takane, de Leeuw 1987; Bartholomew 1987).
  - Wyraźnie odmienny w porównaniu do analizy czynnikowej zestaw technik statystycznych używanych do estymacji modeli.
    - Znikome przenikanie się tradycji KTT i IRT w praktyce badawczej.

Model psychometryczny

Zakładamy, że:

- Cechy, które chcemy mierzyć nie są obserwowalne bezpośrednio (są to cechy ukryte), lecz jedynie za pośrednictwem przejawów (np. rozwiązań zadań), które pozostają z nimi w zależnościach statystycznych.
  - Postać zależności musimy założyć w modelu, a parametry ją opisujące zwykle estymujemy na podstawie danych.
- Obserwowane związki statystyczne pomiędzy różnymi przejawami (wynikami zadań) wynikają z tego (i tylko z tego), że są one przejawami tej samej cechy (zgodnie ze specyfikacją modelu).
  - To założenie nie zawsze musi się sprawdzać – model psychometryczny jako hipoteza.
- Zmienne opisujące przejawy badanych cech zwykle nie są ciągłe, lecz mają charakter porządkowy (i to o niewielkiej liczbie różnych przyjmowanych wartości), w szczególności mogą to być zmienne dychotomiczne (0-1).

Modele psychometryczne

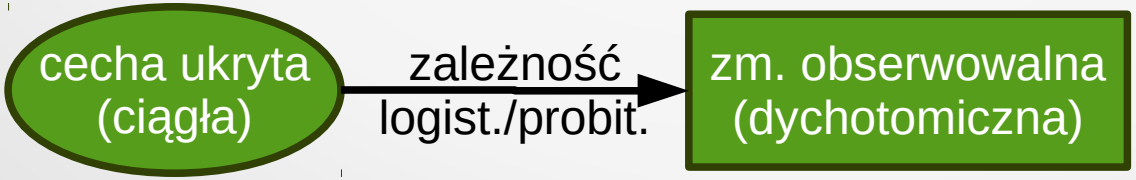
Modele IRT i równoważne oraz pokrewne – ogólna klasyfikacja:

<table>
<tr><th rowspan="2">mierzone cechy są zmiennymi</th><th colspan="2">zmienne opisujące przejawy</th><th rowspan="2">uwagi</th></tr>
<tr><th>dychotomiczne</th><th>porządkowe</th></tr>
<tr><td rowspan="3">ciągłymi
(określone na ℝ;
na potrzeby estymacji bardzo często przyjmuje się, że posiadają one rozkład normalny)</td><td>Rasch, OPLM, 2PL, 3PL, 4PL i ich wersje probitowe</td><td rowspan="2">(S)GRM, RGRM, GPCM, RPCM, RSM i inne oraz ich wersje probitowe</td><td rowspan="3">jednowymiarowe lub wielowymiarowe</td></tr>
<tr><td>nieparametryczne modele IRT (Mokken)</td></tr>
<tr><td colspan="2">modele regresji latentnej, wielopoziomowe modele IRT</td></tr>
<tr><td>porządkowymi
(o niewielkiej liczbie przyjmowanych wartości)</td><td colspan="2" rowspan="2">Cognitive Diagnostic Models</td><td rowspan="2">co do zasady wielowymiarowe</td></tr>
<tr><td>dychotomicznymi</td></tr>
</table>

Dwie tradycje

- **Konfirmacyjna analiza czynnikowa** (CFA) i SEM, estymowane z macierzy korelacji między zmiennymi.
  - Dostosowanie założeń modelu do porządkowego/ dychotomicznego charakteru zmiennych – estymacja z macierzy korelacji polichorycznych.

- **Item Response Theory** – od początku konstruowana z myślą o tym, że przejawy zmiennych ukrytych mierzone są na skalach porządkowych Estymacja z pełnej macierzy danych.



Oba podejścia bardzo wiele łączy, a pod pewnymi warunkami są wręcz formalnie równoważne!

Dwie tradycje

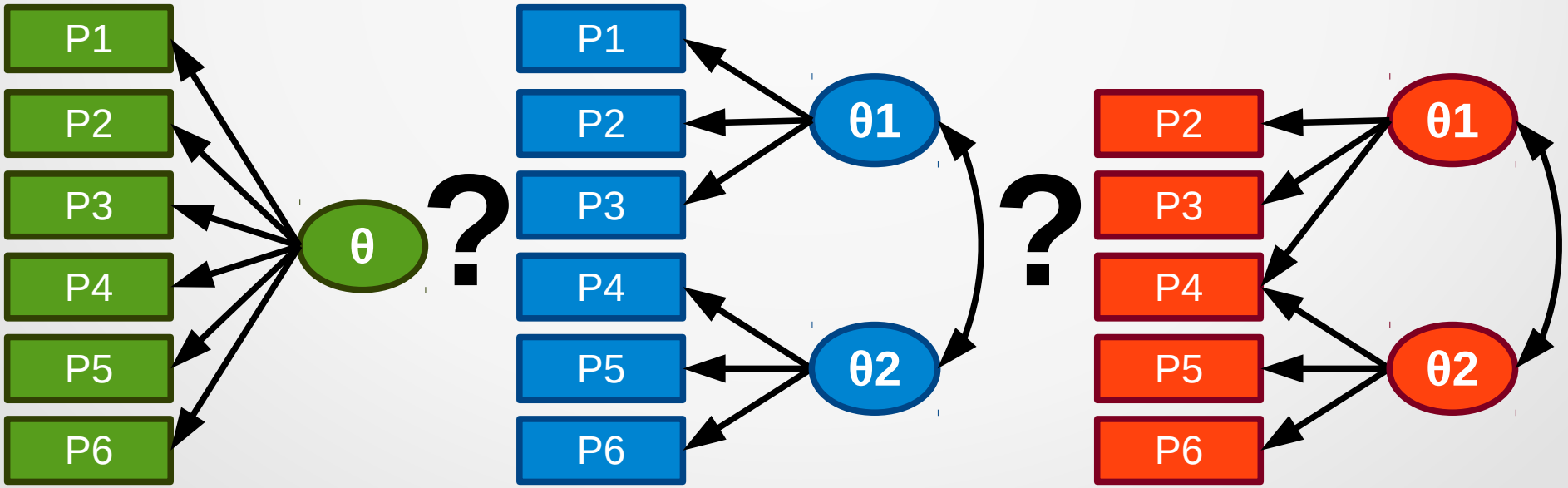
- **Konfirmacyjna analiza czynnikowa / SEM estymowane z macierzy korelacji:**
  - Podejście mniej złożone obliczeniowo.
  - Cała masa indeksów pozwalających oceniać jakość dopasowania modelu do danych.
  - Równoważna założeniu o probitowej funkcji łączącej.
  - Nie pozwala uwzględnić modeli 3PL i 4PL.
  - Nie da się zastosować do typowych schematów badawczych z planowymi brakami danych (musi dać się wyliczyć korelację między każdą parą zmiennych w modelu).

Dwie tradycje

- **Item Response Theory:**
  - Większa swoboda wyboru modelu.
  - Da się zastosować do schematów badawczych z planowymi brakami danych (oczywiście musi występować pewna pula pytań wspólnych).
  - Ładnie ilustruje się wykresami.
  - Złożona obliczeniowo, zwłaszcza dla modeli wielowymiarowych (całkowanie numeryczne).
  - Problemy z oceną jakości dopasowania modelu do danych (najlepiej symulacyjnie, ale jest to możliwe tylko dla relatywnie prostych modeli).

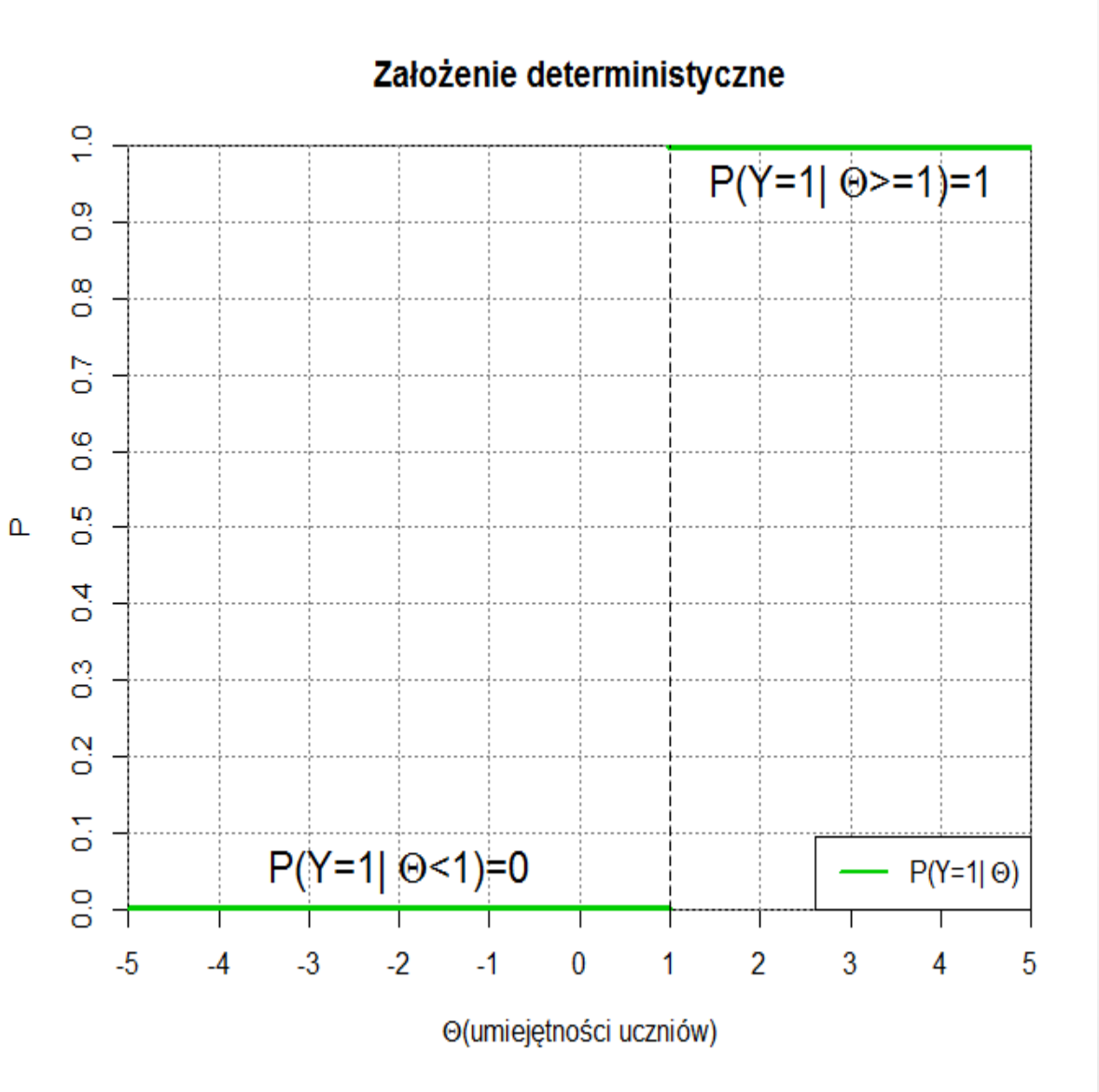
Model psychometryczny - hipoteza

- Czy są podstawy by twierdzić, że dany zestaw pytań mierzy (w pewnym sensie) tę samą cechę?
- Jak mógłby wyglądać model, który lepiej opisywałbym (potencjalne) przyczyny obserwowanych zależności?



Modelowanie zależności pomiędzy mierzoną cechą a wynikiem rozwiązania zadania

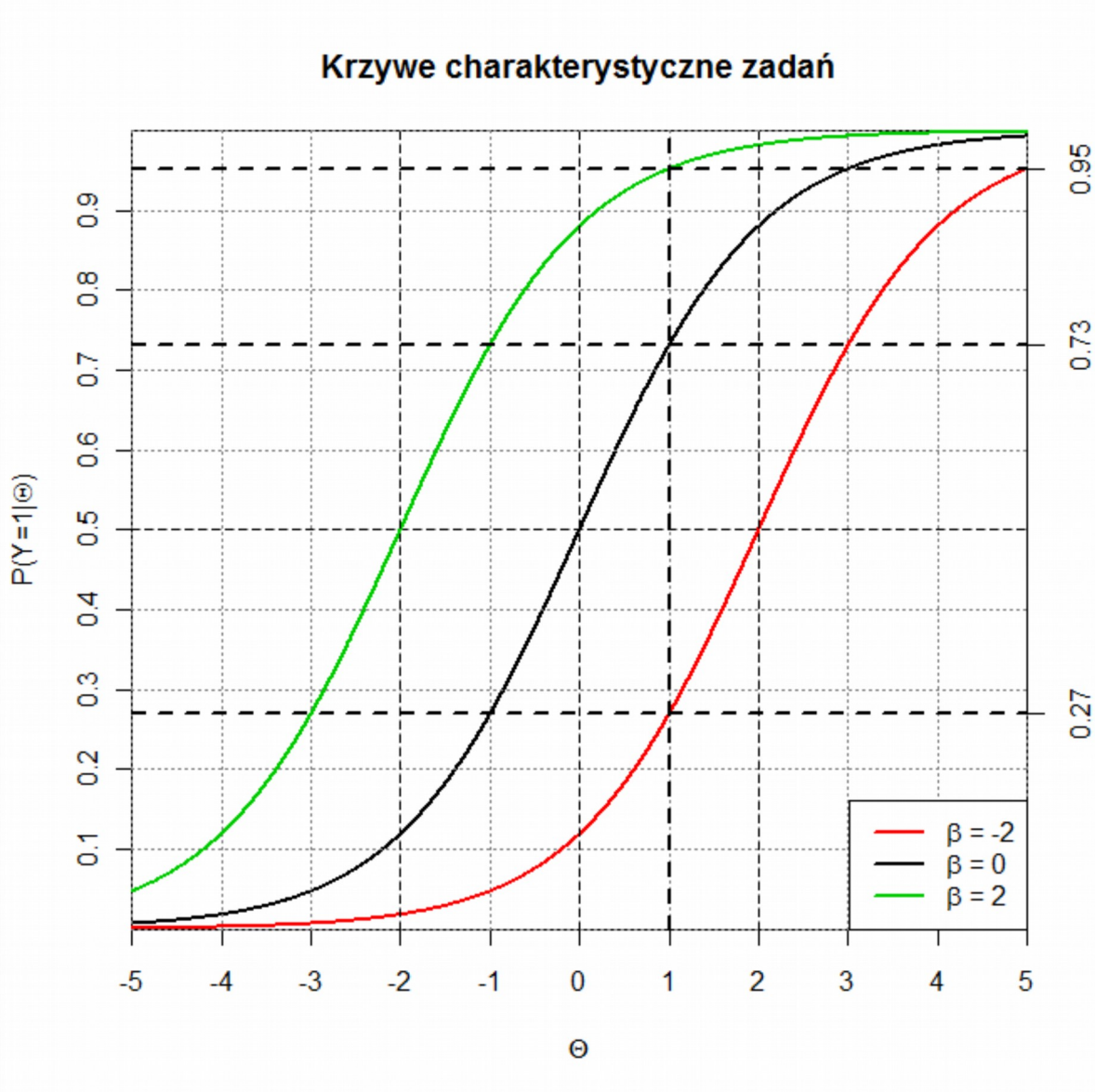
Skalogram Guttmana (to nie IRT)



- Każde zadanie ma trudność, działającą jako próg.
- Uczniowie o umiejętnościach poniżej progu rozwiązują zadanie niepoprawnie.
- Uczniowie o umiejętnościach powyżej progu rozwiązują zadanie poprawnie.

Na podstawie takich zadań moglibyśmy okreslić tylko uporządkowanie badanych, nie moglibyśmy jednak oszacować wartości cechy na skali przedziałowej!

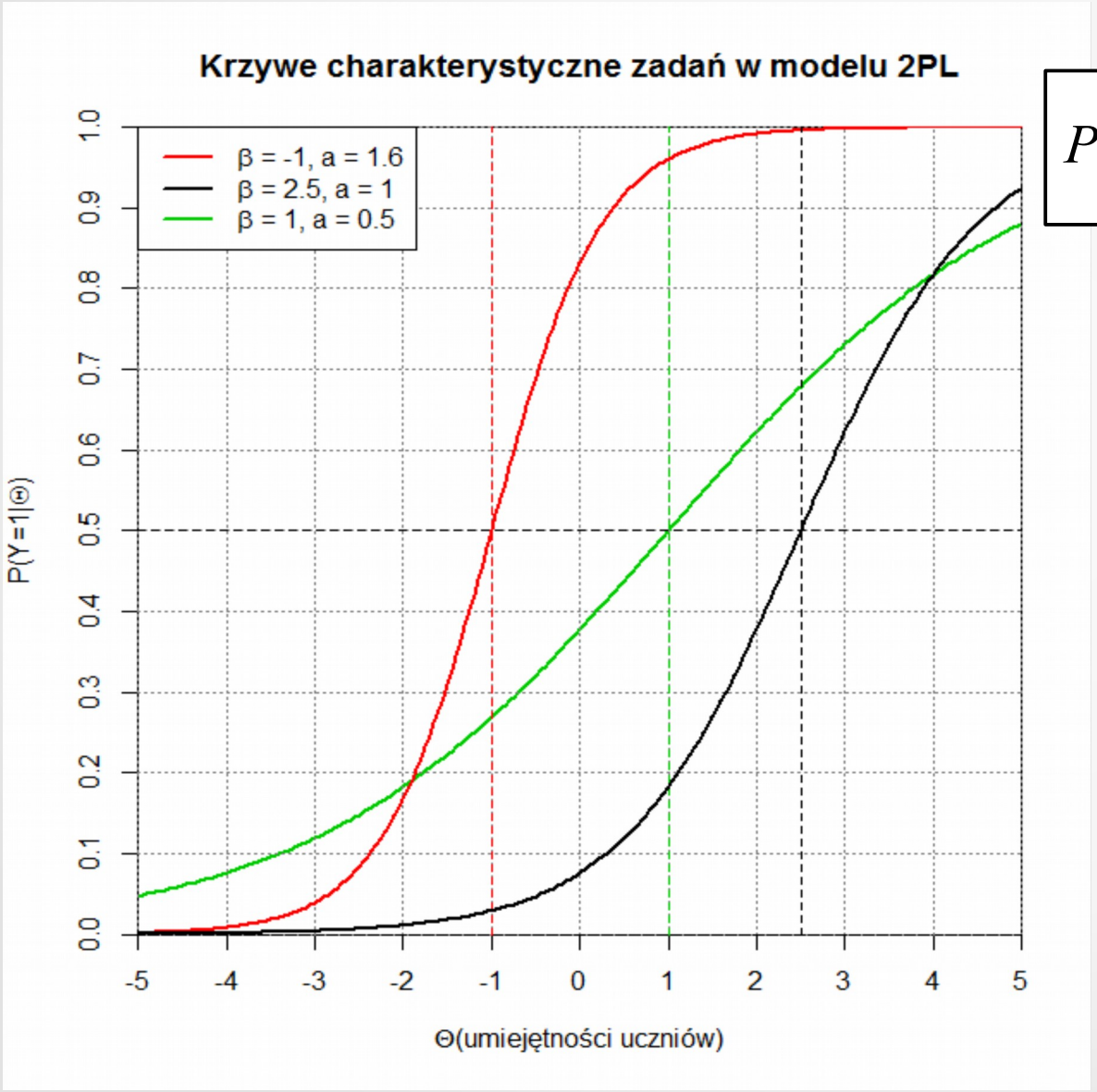
Model Rascha



$$P(X_i = 1|\Theta) = \frac{\exp(\Theta - b_i)}{1 + \exp(\Theta - b_i)}$$

- Prawdopodobieństwa udzielenia poprawnej odpowiedzi przez ucznia o poziomie umiejętności Θ=1 na pytania o **trudności**:
  - β_i=-2 jest równe 0,95;
  - β_i= 0 jest równe 0,73;
  - β_i= 2 jest równe 0,27.
- **Trudność** pytania to poziom umiejętności, dla którego prawd. poprawnej odpowiedzi jest równe 0,5

Model 2PL

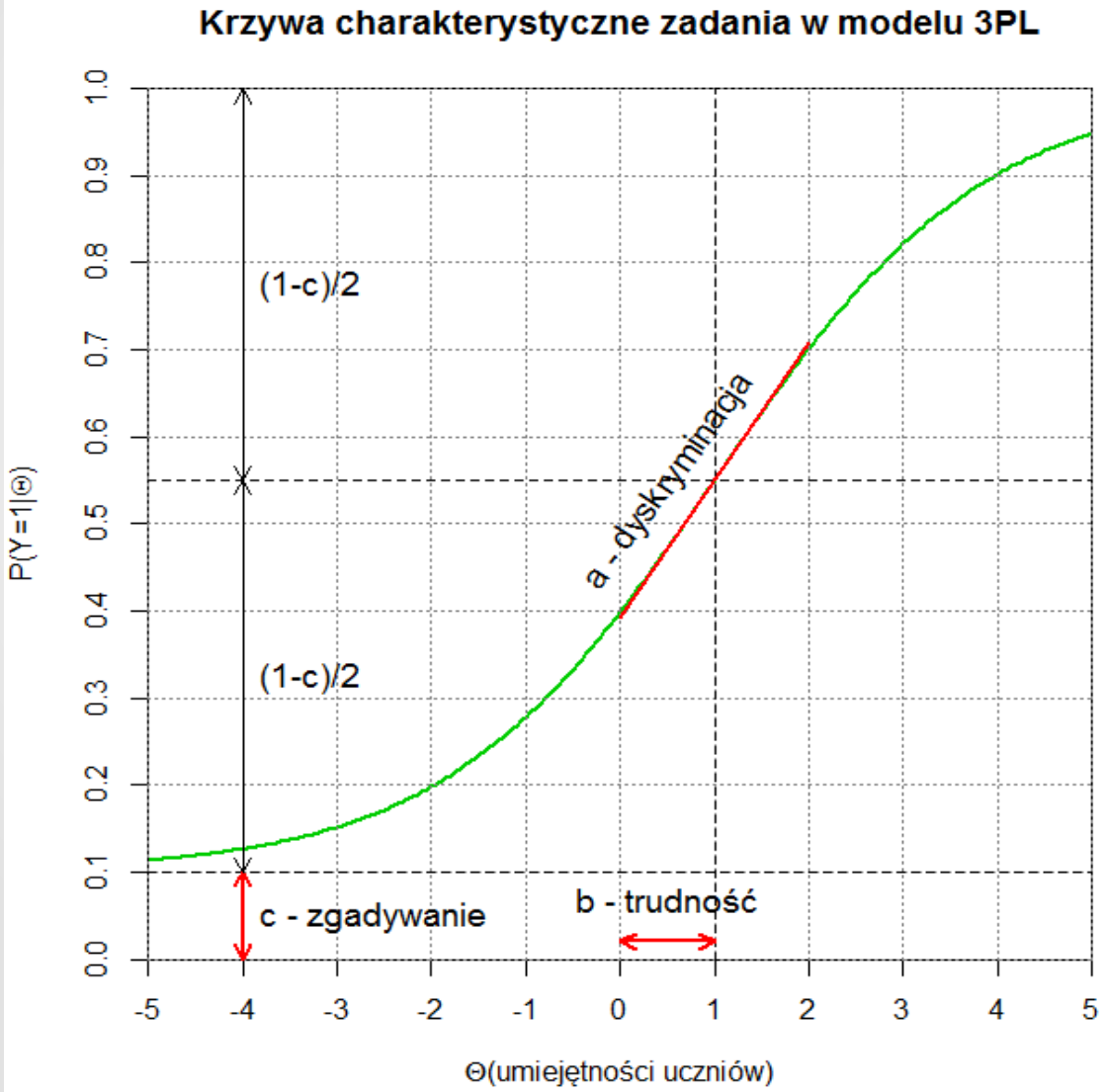


$$P(X_i = 1|\Theta) = \frac{\exp[a_i(\Theta - b_i)]}{1 + \exp[a_i(\Theta - b_i)]}$$

- Wartość parametru **dyskryminacji** (a_i) wpływa na nachylenie krzywej charakterystycznej zadania.
- Im wyższa dyskryminacja, tym bardziej odpowiedź na dane pytanie związana z mierzoną cechą.
- **Trudność** pytania (β_i) przesuwa krzywą charakterystyczną w poziomie.
- Krzywe mogą się przecinać.

Model 3PL

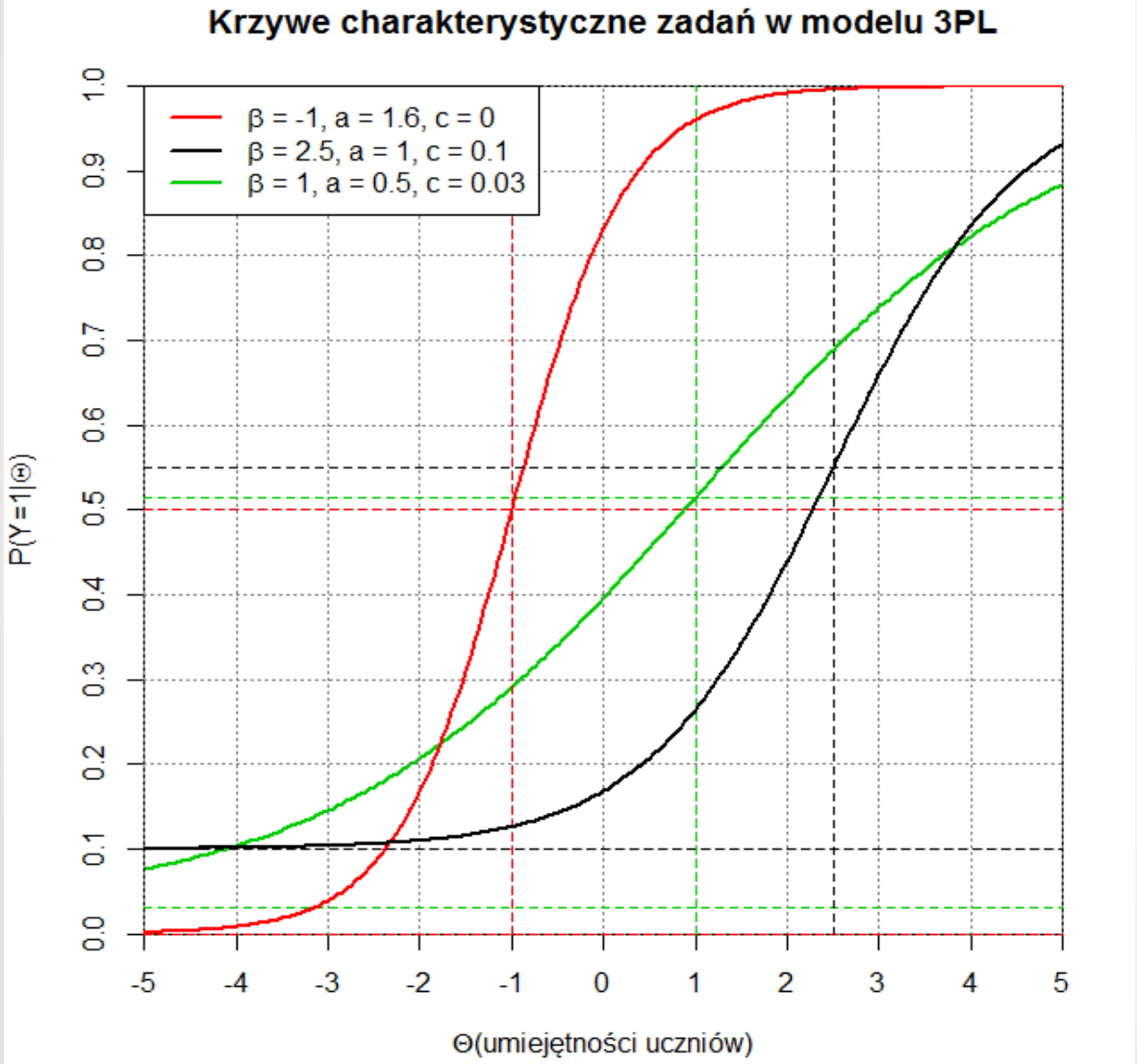
$$P(Y_{ij}=1|\Theta_j)=c_i+(1-c_i)\frac{\exp(a_i(\Theta_j-\beta_i))}{1+\exp(a_i(\Theta_j-\beta_i))}$$



- Wartość parametru (pseudo)zgadywania (c_i) *podnosi* poziom, do którego zbiega krzywa charakterystyczna zadania dla bardzo niskich umiejętności.
- Chcemy, by była ona jak najmniejsza (bliska 0).
- Wartość parametru dyskryminacji (a_i) wpływa na nachylenie krzywej charakterystycznej zadania.
- Trudność pytania (β_i) przesuwa krzywą charakterystyczną w poziomie.

Model 3PL

$$P(Y_{ij}=1|\Theta_j)=c_i+(1-c_i)\frac{\exp(a_i(\Theta_j-\beta_i))}{1+\exp(a_i(\Theta_j-\beta_i))}$$



- Wartość parametru (pseudo)zgadywania (c_i) *podnosi* poziom, do którego zbiega krzywa charakterystyczna zadania dla bardzo niskich umiejętności.
- Chcemy, by była ona jak najmniejsza (bliska 0).
- Wartość parametru dyskryminacji (a_i) wpływa na nachylenie krzywej charakterystycznej zadania.
- Trudność pytania (β_i) przesuwa krzywą charakterystyczną w poziomie.

Model Rascha – wady i zalety

Zalety:

- Elegancki formalnie, nieproblematyczny w interpretacji.
  - Zadania jednoznacznie uszeregowane ze względu na trudność.
  - Każde zadanie ma taki sam wkład w przewidywany poziom umiejętności uczniów (pozwala odwzorowaćw przewidywaniach umiejętności uczniów założone proporcje treściowe/dziedzinowe).
- Relatywnie łatwy w estymacji (mała liczba parametrów).

Wady:

- Trudno ułożyć test, zwłaszcza o szerokim zakresie treściowym, który będzie dobrze pasował do założeń modelu Rascha.
- Mało użyteczny na etapie badań pilotażowych (oceny jakości zadań).

Model 3PL – wady i zalety

Zalety:

- Pozwala modelować zgadywanie – czynnik, który wydaje się ważnym problemem w testach badających wiedzę i umiejętności.
- Bardzo użyteczny w badaniach pilotażowych – daje dużo informacji o sposobie funkcjonowania zadania.

Wady:

- Trudny w estymacji (duża liczba parametrów).
  - W przypadku łatwych zadań trudności z wiarygodnym oszacowaniem wartości parametrów zgadywania.
- Możliwość występowania znacznych różnic we wkładzie poszczególnych zadań w przewidywane umiejętności uczniów.
  - Jeśli .pytania dotyczące jakichś dziedzin będą miały gorsze własności pomiarowe, dziedzina ta będzie niedoreprezentowana w uzyskanych przewidywaniach umiejętności uczniów.

Model 2PL – wady i zalety

Zalety:

- Nie sprawia większych trudności w estymacji.
- Większość testów w zadowalający sposób spełnia jego założenia.
- Nawet jeśli występuje zgadywanie, oszacowania poziomu umiejętności uczniów z modelu 2PL są bardzo zbliżone do tych, jakie uzyskałoby się stosując model 3PL
  - Ale już oszacowania błędów standardowych z obu modeli będą wyraźnie się różnić.

Wady:

- Możliwość występowania znacznych różnic we wkładzie poszczególnych zadań w przewidywane umiejętności uczniów.
  - Jeśli pytania dotyczące jakichś dziedzin będą miały gorsze własności pomiarowe, dziedzina ta będzie niedoreprezentowana w uzyskanych przewidywaniach umiejętności uczniów

Dwa słowa o estymacji

Metody estymacji

Tradycja CFA (estymacja z macierzy korelacji):

1. Wyestymuj macierz korelacji polichorycznych.
2. Na jej podstawie wyestymuj parametry modelu.
   - Preferowana metoda Weighted Least Squares – musimy brać pod uwagę ten problem, że wariancja zmiennych obserwowalnych jest powiązana z ich średnią.
3. Ewentualnie wylicz oszacowania wartości cech ukrytych.
   - Wiele możliwych metod.

Metody estymacji

Tradycja IRT (estymacja z pełnej macierzy danych):

Modele Rascha i OPLM:

- Wiele metod: Joint ML, Conditional ML (obie nie nakładają założeń na rozkład badanej cechy), Marginal ML, metody bayesowskie.

Bardziej złożone modele:

- Marginal ML – zakładamy rozkład badanej cechy w populacji, z której pochodzi badana grupa; metody bayesowskie.

1. Wyestymuj parametry modelu.
2. Ewentualnie wylicz oszacowania wartości cech ukrytych.
   - Wiele możliwych metod.

Przewidywanie poziomu cechy

- Poziom umiejętności przewidywany dla ucznia na podstawie punktacji, jaką uzyskał on z testu zależy od własności pomiarowych (jakości) zadań, które uczeń rozwiązał poprawnie:
  - Liczby zadań, które rozwiązał poprawnie – w modelu Rascha.
  - Parametrów dyskryminacji zadań, które rozwiązał poprawnie – w modelu 2PL.
  - Parametrów dyskryminacji i zgadywania zadań, które rozwiązał poprawnie – w modelu 3PL.
- Przewidywany poziom umiejętności **nie** zależy od trudności zadań, które uczeń rozwiązał poprawnie (jeśli tylko wszyscy zdający rozwiązywali ten sam zestaw zadań).

Przewidywanie poziomu cechy

Błąd oszacowania umiejętności różni się dla poszczególnych badanych. Aby go określić stosowane jest jedno z dwóch podejść:

- Na podstawie tzw. krzywej informatycznej testu (błąd pomiaru jest funkcją wartości mierzonej cechy).
  - Podejście powiązane z estymatorami poziomu cechy ML i WML.
  - Ogólnie rzecz biorąc, zadanie daje dużo informacji w zakresie umiejętności zbliżonych do swojego poziomu trudności.

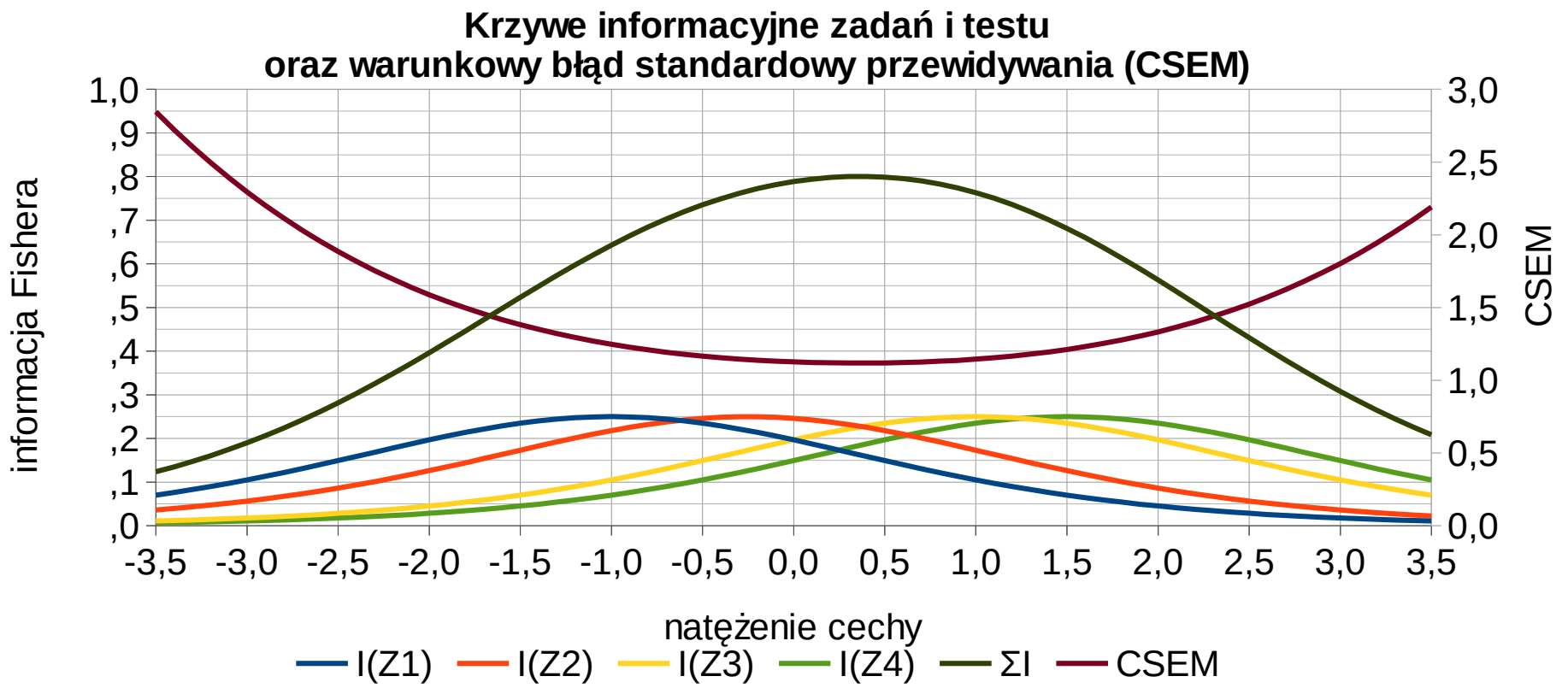
$$I_i(\theta)=\left[a_i\frac{P\left(X_i=1|\theta\right)-c_i}{1-c_i}\right]^2\frac{P\left(X_i=0|\theta\right)}{P\left(X_i=1|\theta\right)}$$

- Na podstawie odchylenia standardowego rozkładu warunkowego wartości mierzonej cechy ze względu na profil odpowiedzi na zadania (błąd pomiaru jest funkcją profilu odpowiedzi).
  - Podejście powiązane z estymatorem poziomu cechy EAP.

Przewidywanie poziomu cechy

Przykładowa krzywa informacyjna i błąd standardowy przewidywania dla testu złożonego z czterech zadań, w modelu Rascha:

- Trudności zadań: $b_1 = -1$, $b_2 = -0{,}25$, $b_3 = 1$, $b_4 = 1{,}5$, dyskryminacja 1.



Problemy z dychotomicznymi (i porządkowymi) zmiennymi obserwowanymi

Problemy: rozkład sumy punktów

- Jeśli odpowiedzi na pytania są przejawami tej samej cechy ukrytej, to najprostszym wskaźnikiem natężenia cechy może być suma punktów przypisanych do udzielonej odpowiedzi.
- Przyzwyczajenie wyniesione z modeli *stricte* liniowych każe nam oczekiwać, że rozkład takiej sumy powinien być (przy dużej liczbie badanych) zbliżony do rozkładu cechy ukrytej...
- … ale gdy cechę ukrytą ze zmiennymi obserwowalnymi łączy zależność logistyczna/probitowa będzie tak tylko pod warunkiem, że dobrze dobraliśmy trudność zadań.

Problemy: rozkład sumy punktów

Mała symulacja:

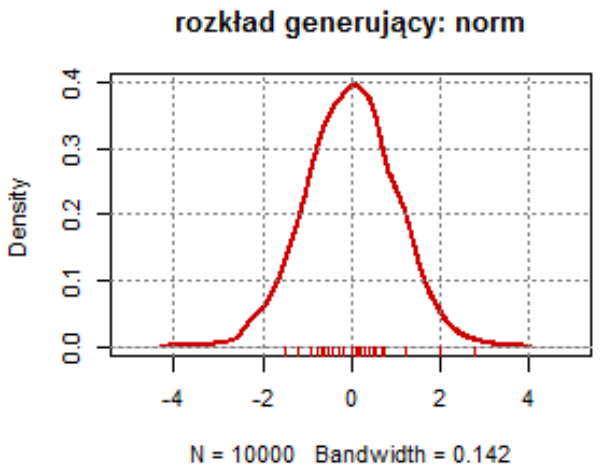
- Wygenerujmy trzy duże grupy *badanych*, losując im wartości cechy spod trzech różnych rozkładów: normalnego, log-normalnego i jednostajnego (a następnie wystandaryzujmy wartości cechy w każdej grupie do średniej 0 i odch. stand. 1).
- Dajmy im do *rozwiązania* pięć różnych *testów*, różniących się trudnością pytań. Dla uproszczenia załóżmy, że *odpowiadają* na nie zgodnie z założeniami modelu Rascha. Każdy *test* składa się z 30 pytań o trudnościach wylosowanych z:

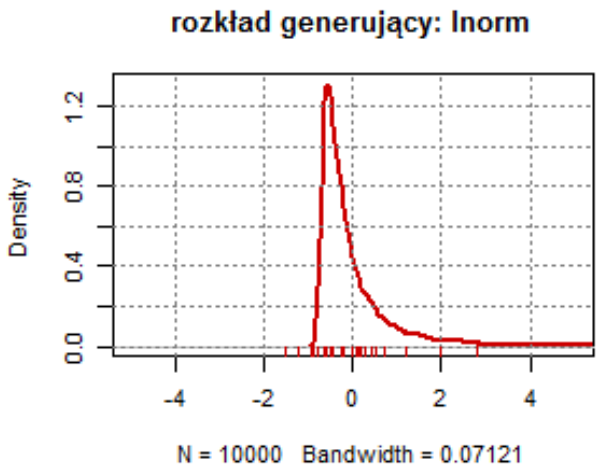
  1) N(0, 1) 2) mieszaniny 1:1 N(-1.5, 0.5) i N(1.5, 0.5),

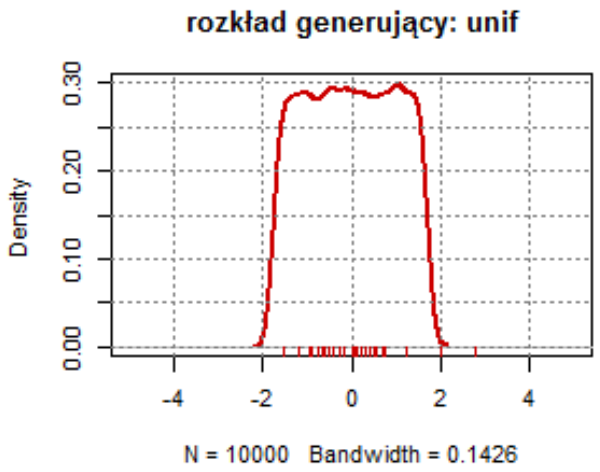
  3) N(0, 0.25) 4) N(1.5, 0.5) 5) N(-1.5, 0.5)
- Sprawdźmy, jak będą wyglądały rozkłady sum punktów.

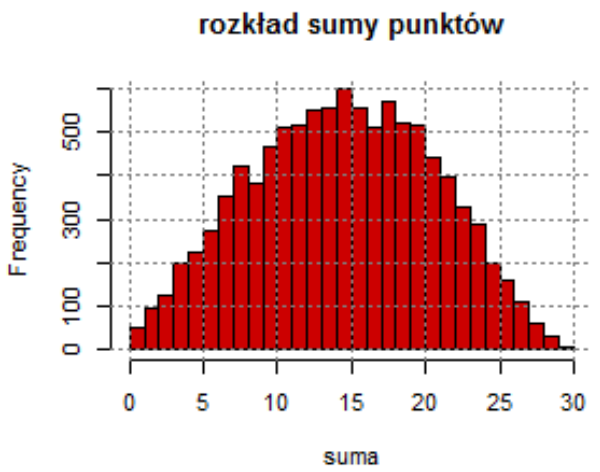
Dosyć dobrze dobrane trudności

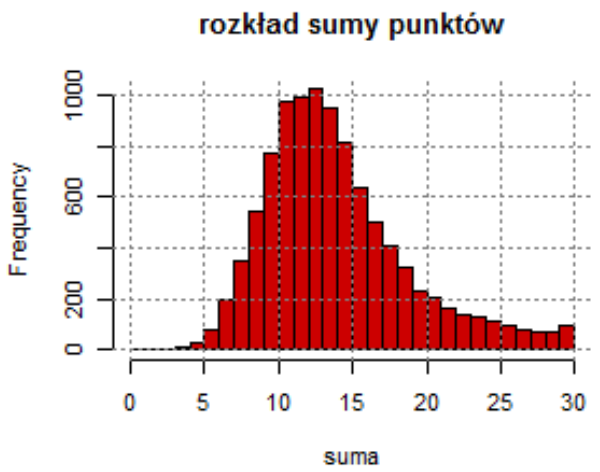
kształty rozkł. sumy punktów zbliżone do rozkł. generujących

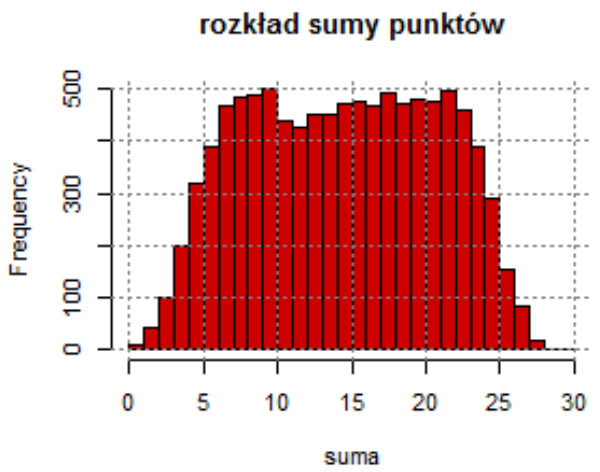






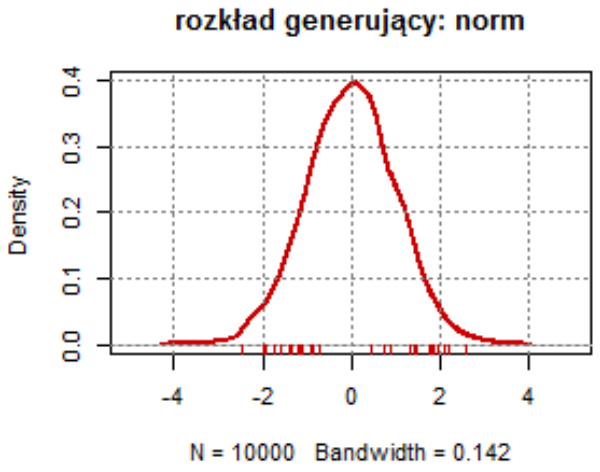


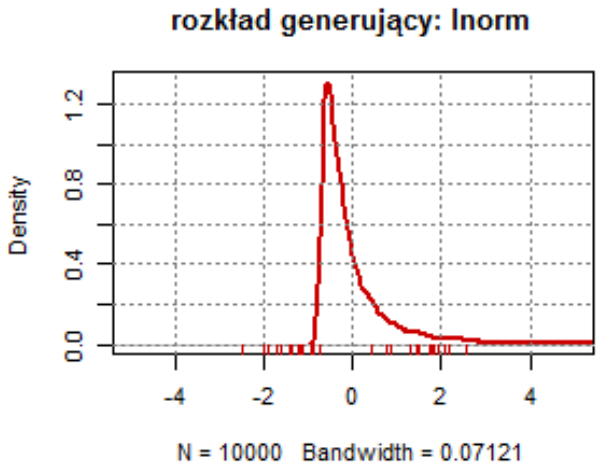


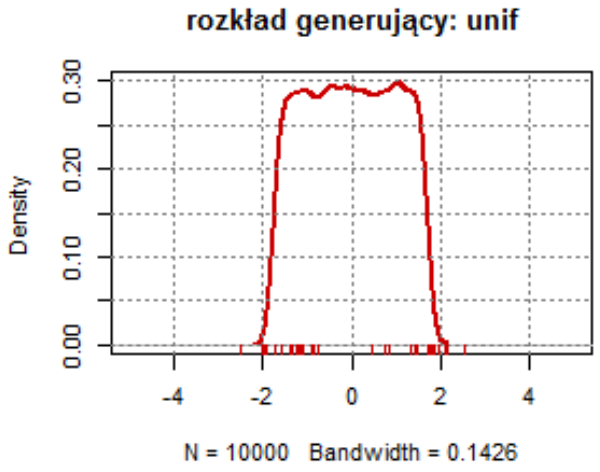


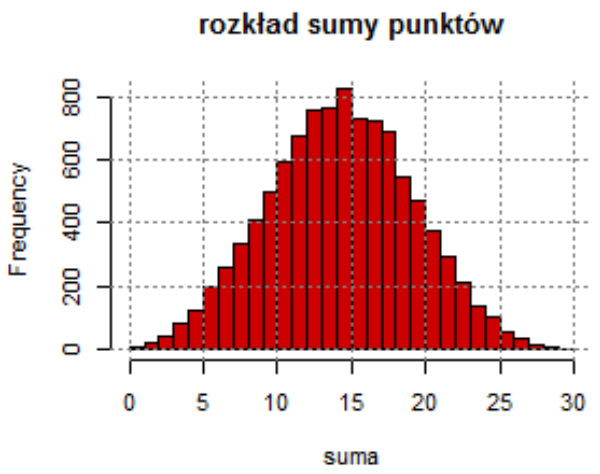
Brak zadań o średniej trudności

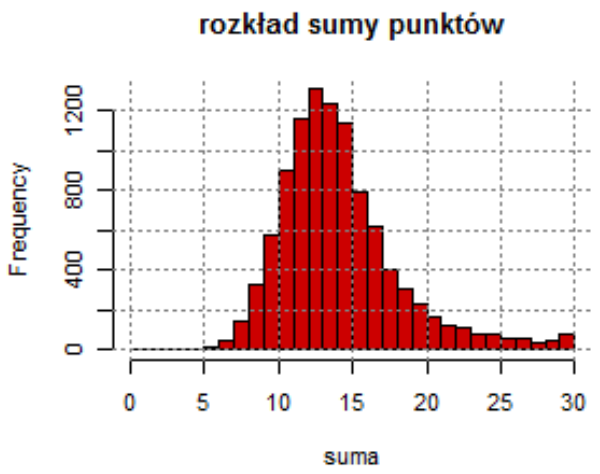
kształty rozkł. sumy punktów zbliżone do rozkł. generujących

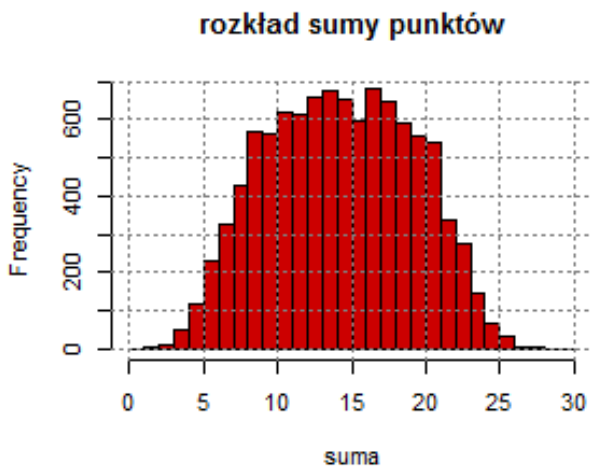






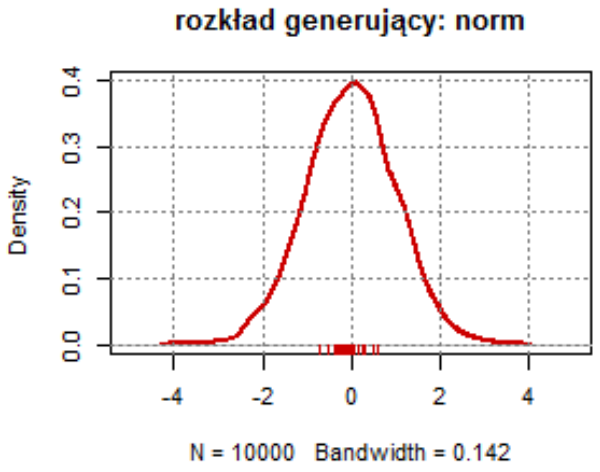


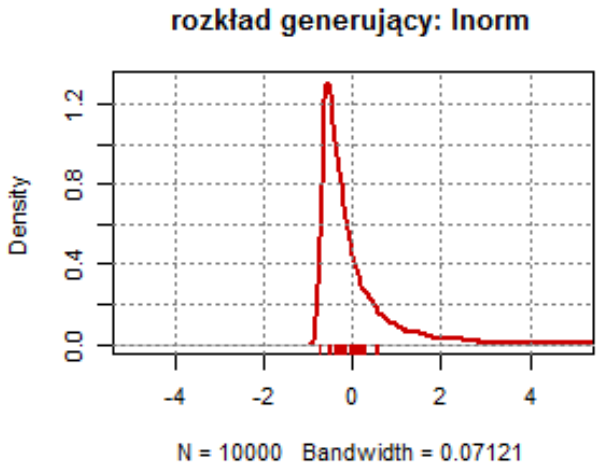


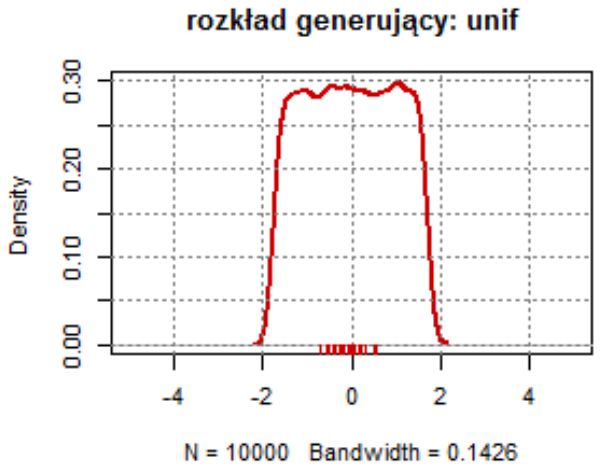


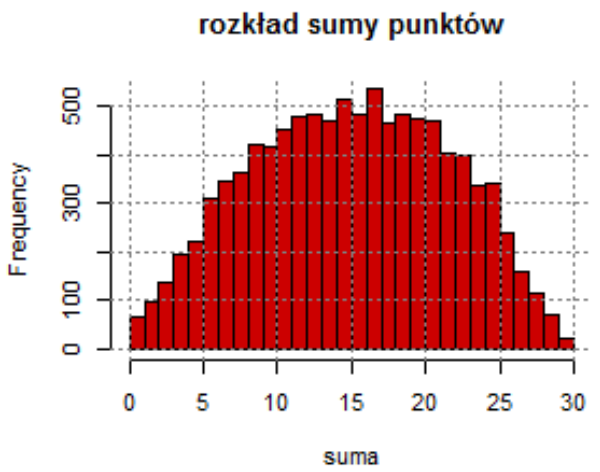
Zbyt małe zróżnicowanie trudności

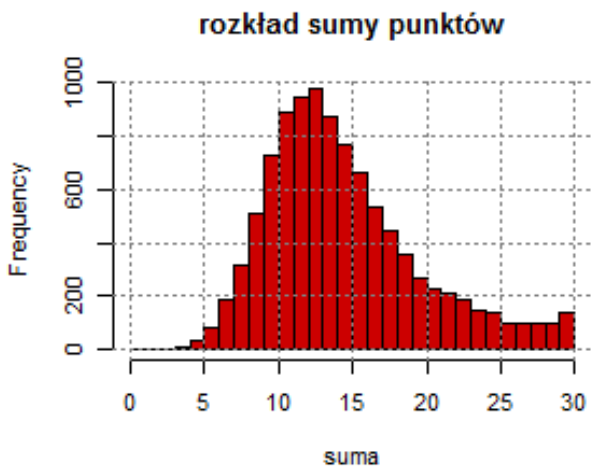
„rozszerzenie” rozkł. sumy punktów w jego środkowej części

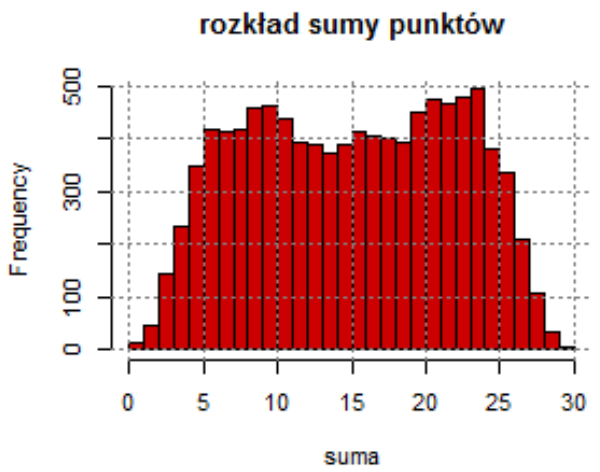






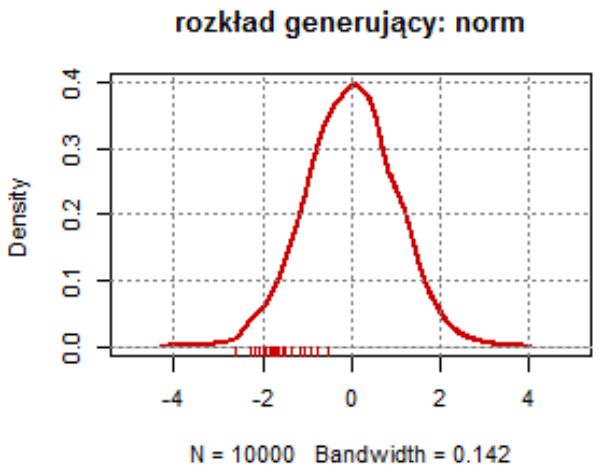


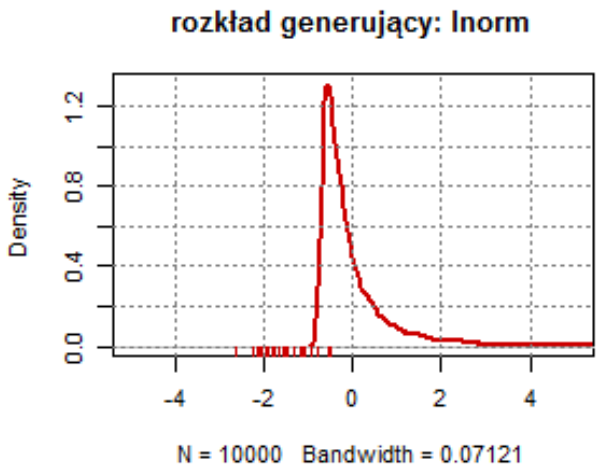


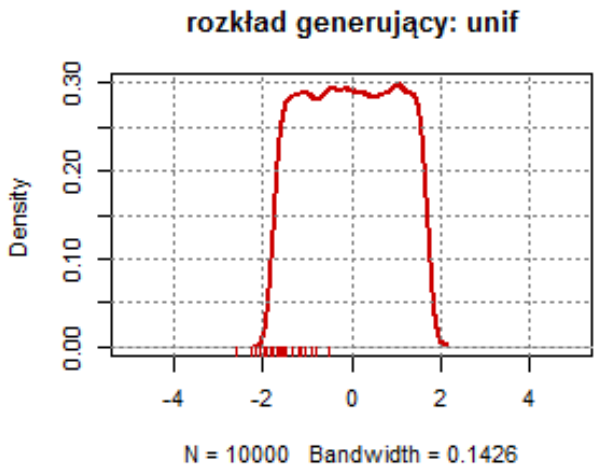


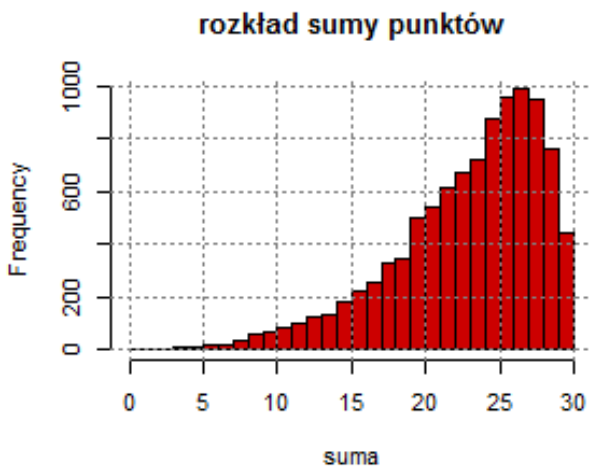
Brak trudnych zadań

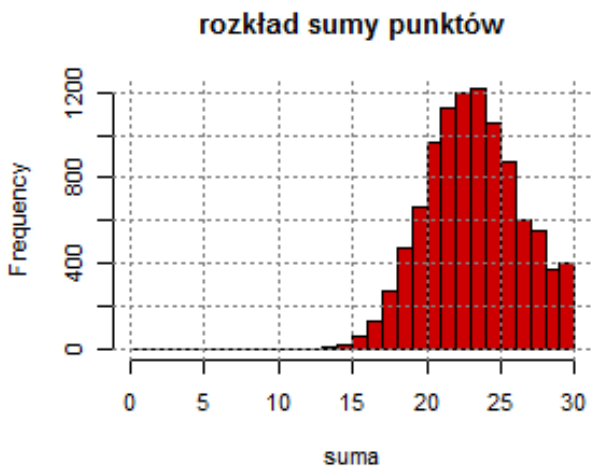
kształty rozkł. sumy punktów są do siebie bardzo podobne, bez względu na rozkł. generujący

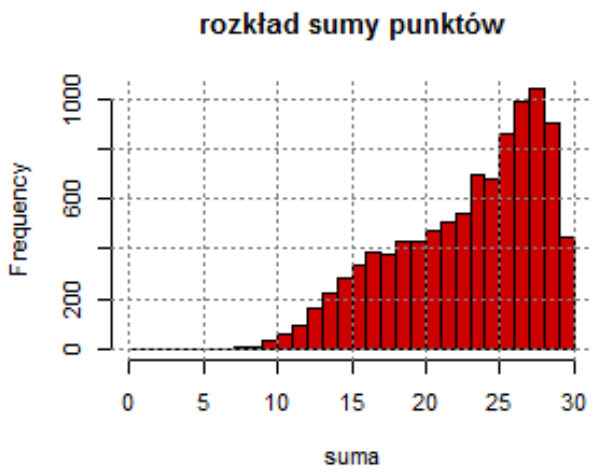






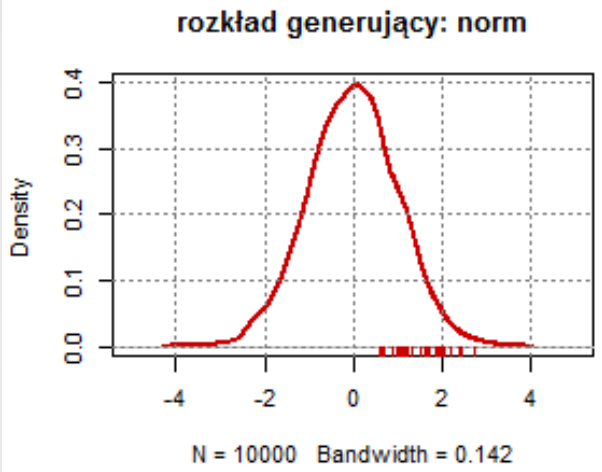


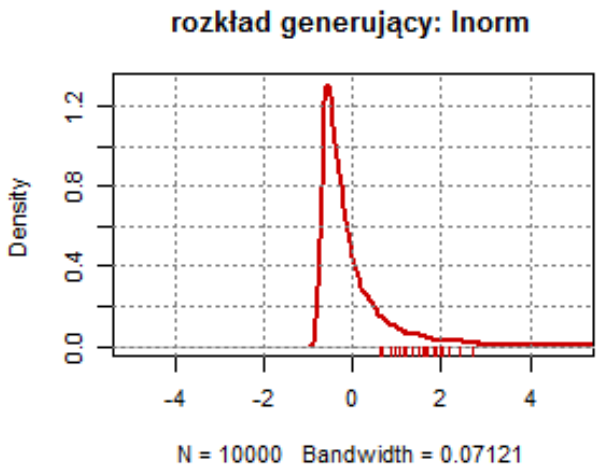


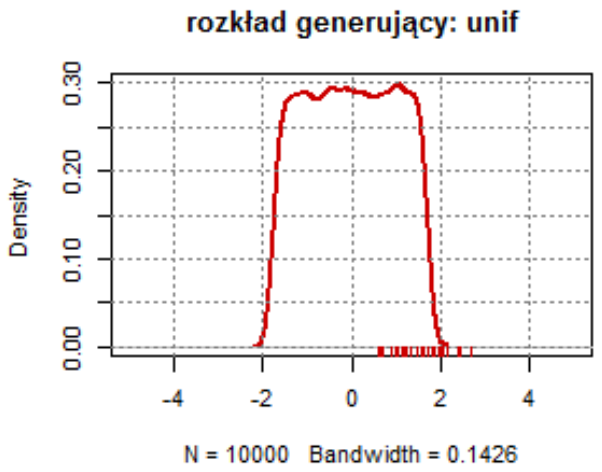


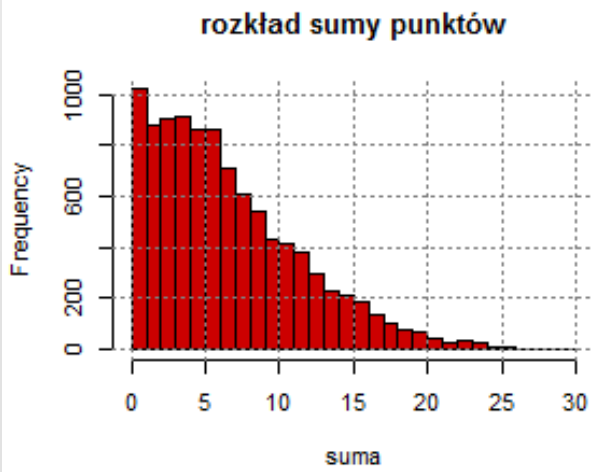
Brak łatwych zadań

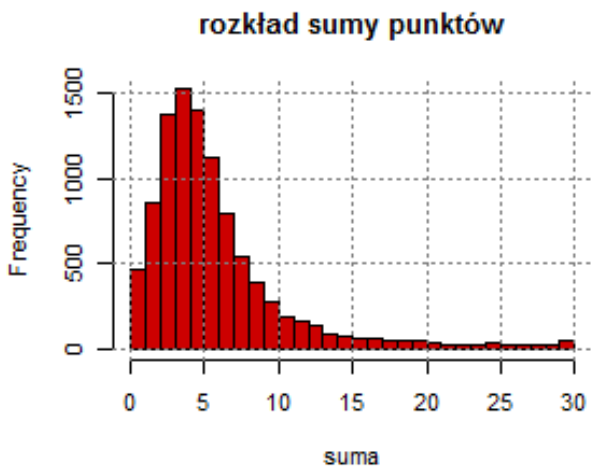
kształty rozkł. sumy punktów są do siebie bardzo podobne, bez względu na rozkł. generujący

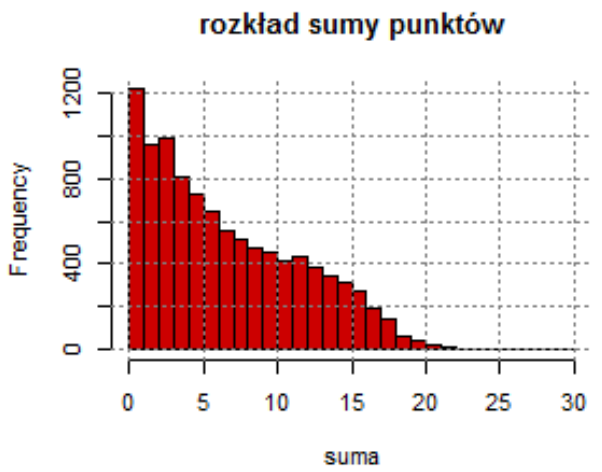












Problemy: skala wyników

- W modelach z ciągłą cechą ukrytą skalę zmiennej ukrytej musimy *zapożyczyć* z którejś zmiennej obserwowanej.
  - Jednak gdy cechę ukrytą ze zmiennymi obserwowalnymi łączy zależność logistyczna/ probitowa, taka skala jest bardzo kłopotliwa interpretacyjnie.
- Alternatywnie możemy ustalić skalę w odniesieniu do przewidywanych parametrów rozkładu zmiennej ukrytej w ramach badanej grupy (lub innej grupy, przebadanej już wcześniej tym samym testem).
  - Jest to rozwiązanie ułatwiające interpretację... ale często nie tak bardzo.

Problemy: skala wyników

Definicja skali PISA:

Wyniki testu PISA określone są na skali takiej, że w roku stanowiącym punkt odniesienia:

- Średnia wyników uczniów z krajów OECD biorących udział w badaniu, wyliczona tak, że każdy kraj ma równy wkład w wyliczaną średnią, jest równa 500.
- Odchylenie standardowe wyników uczniów z krajów OECD biorących udział w badaniu, wyliczona tak, że każdy kraj ma równy wkład w wyliczane odch. stand., jest równe 100.

I dlatego, aby odbiorcy mieli poczucie, że rozumieją, OECD woli *po prostu* mówić, że wyniki określone są na skali od 0 do 1000 punktów, która ma średnią 500...

Problemy: co mierzy zadanie?

Czy te trzy zadania mierzą to samo?

1. Jurek miał dwa żołnierzyki. Na urodziny dostał od Jacka jeszcze dwa. Ile żołnierzyków ma teraz?
2. 2 + 2 = ?
3. Ania miała dwie lalki. Na urodziny dostał od Zosi jeszcze dwie. Ile lalek ma teraz?

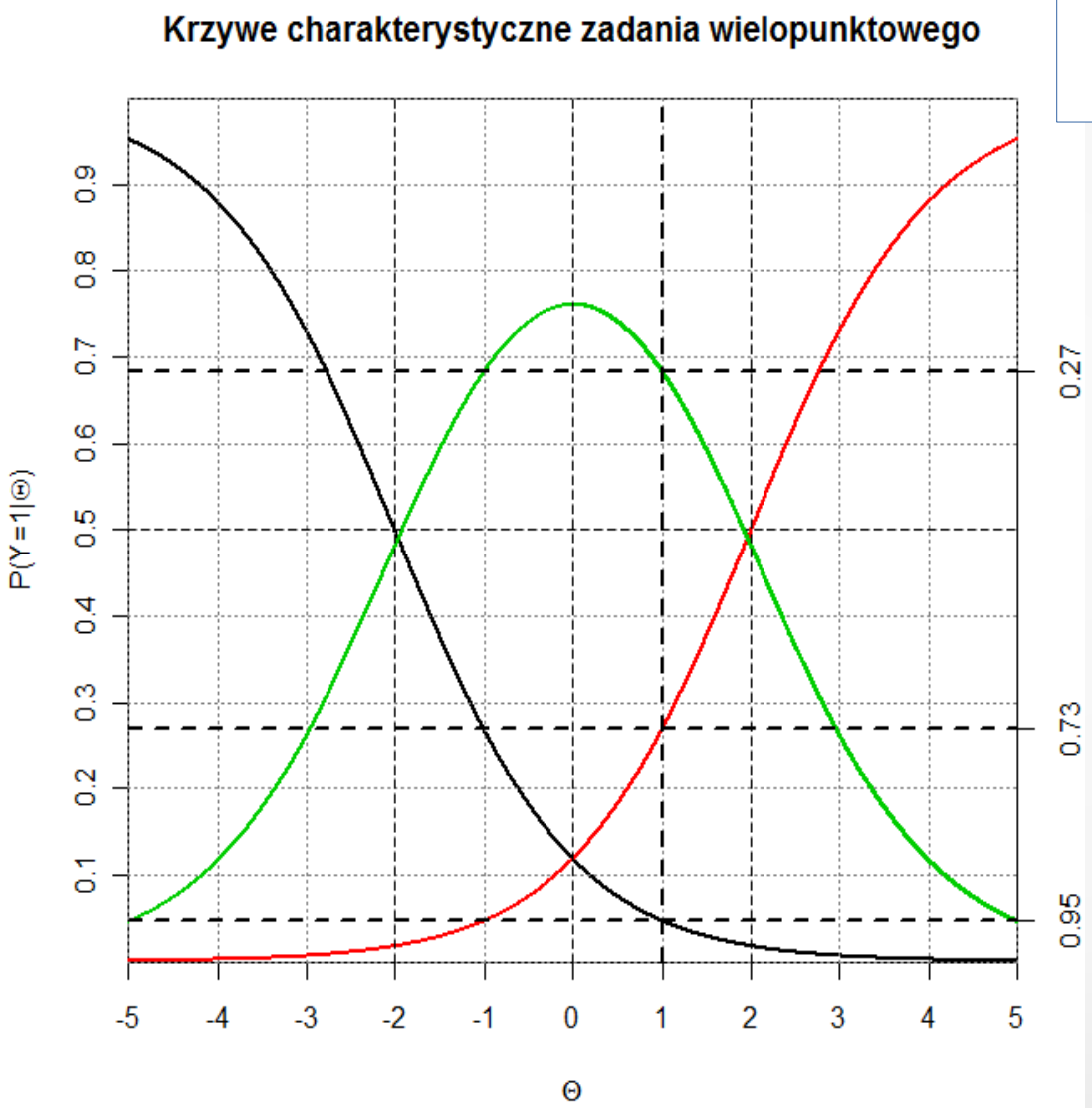
Co do zasady brak nam *obiektywnych* kryteriów oceny, czy zadanie (pytanie) mierzy to, co miało mierzyć. Możemy jednak sprawdzać:

- Czy zadanie mierzy coś podobnego do innych zadań?
- Czy mierzy w ten sam sposób w różnych grupach badanych?

Modelowanie związku badanej cechy z wynikami zadania o wielu poziomach wykonania

Model Partial Credit

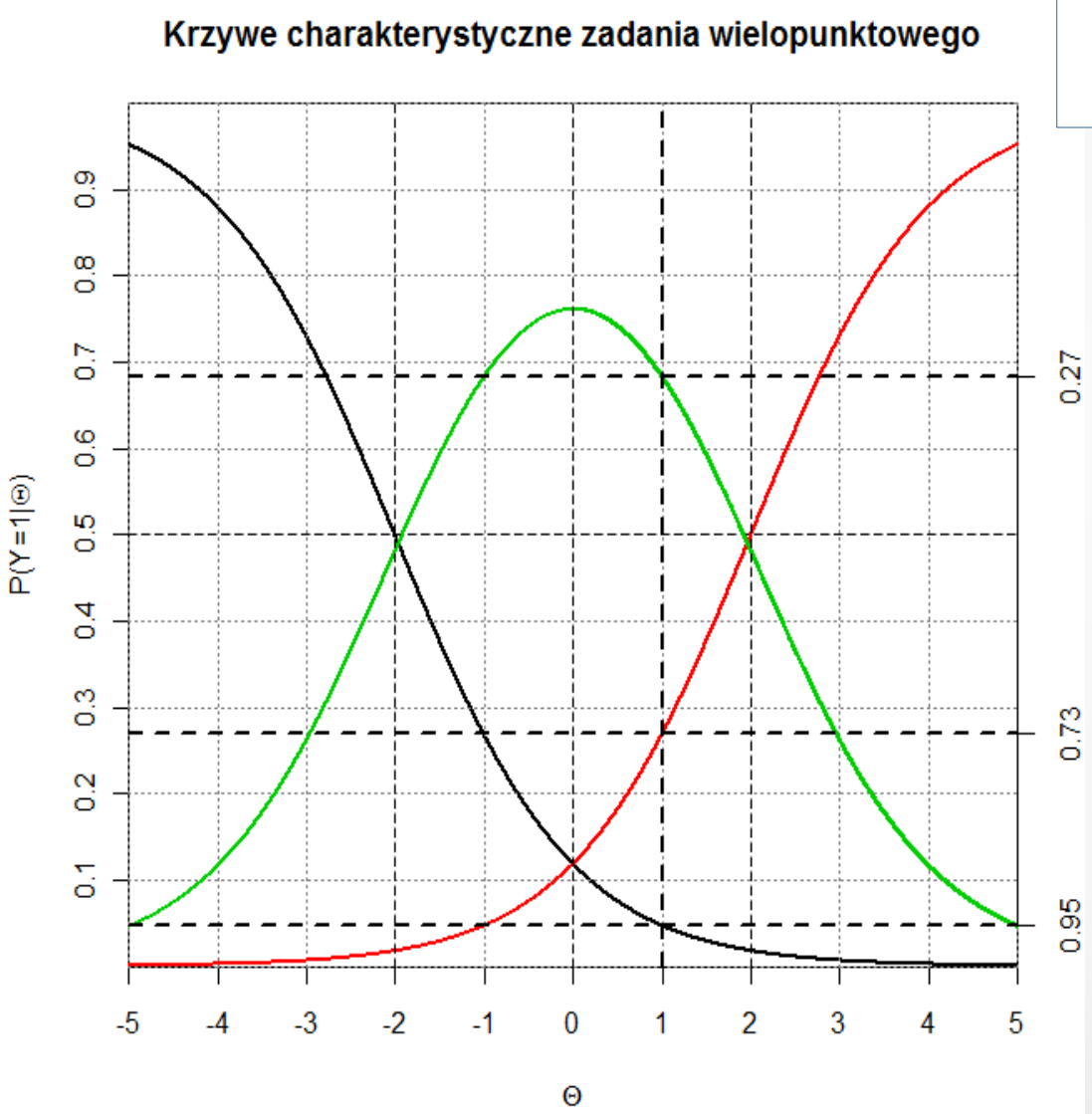
$$P(Y_i = g|\Theta_j) = \frac{\exp\left[\sum_{k=1}^{g} a_i(\Theta - b_{ik})\right]}{1 + \sum_{k=1}^{m_i} P(X_i = k)} \text{ dla } g>0$$



- Krzywe charakterystyczne „środkowych” podpunktów zadania mają inny kształt (dzwonowaty).
- Wartość parametru b_{ik} wskazuje na wartość cechy, dla której następuje przecięcie się krzywych charakterystycznych „kroku” *g* i kroku *g*-1, tj. uzyskanie *g* punktów zaczyna być bardziej prawdopodobne, niż uzyskanie *g*-1 punktów.
- Jako (*ogólną*) trudność zadania można traktować średnią parametrów b_{ik}.

Model Partial Credit

$$P(Y_i = g|\Theta_j) = \frac{\exp\left[\sum_{k=1}^{g} a_i(\Theta - b_{ik})\right]}{1 + \sum_{k=1}^{m_i} P(X_i = k)} \text{ dla } g > 0$$

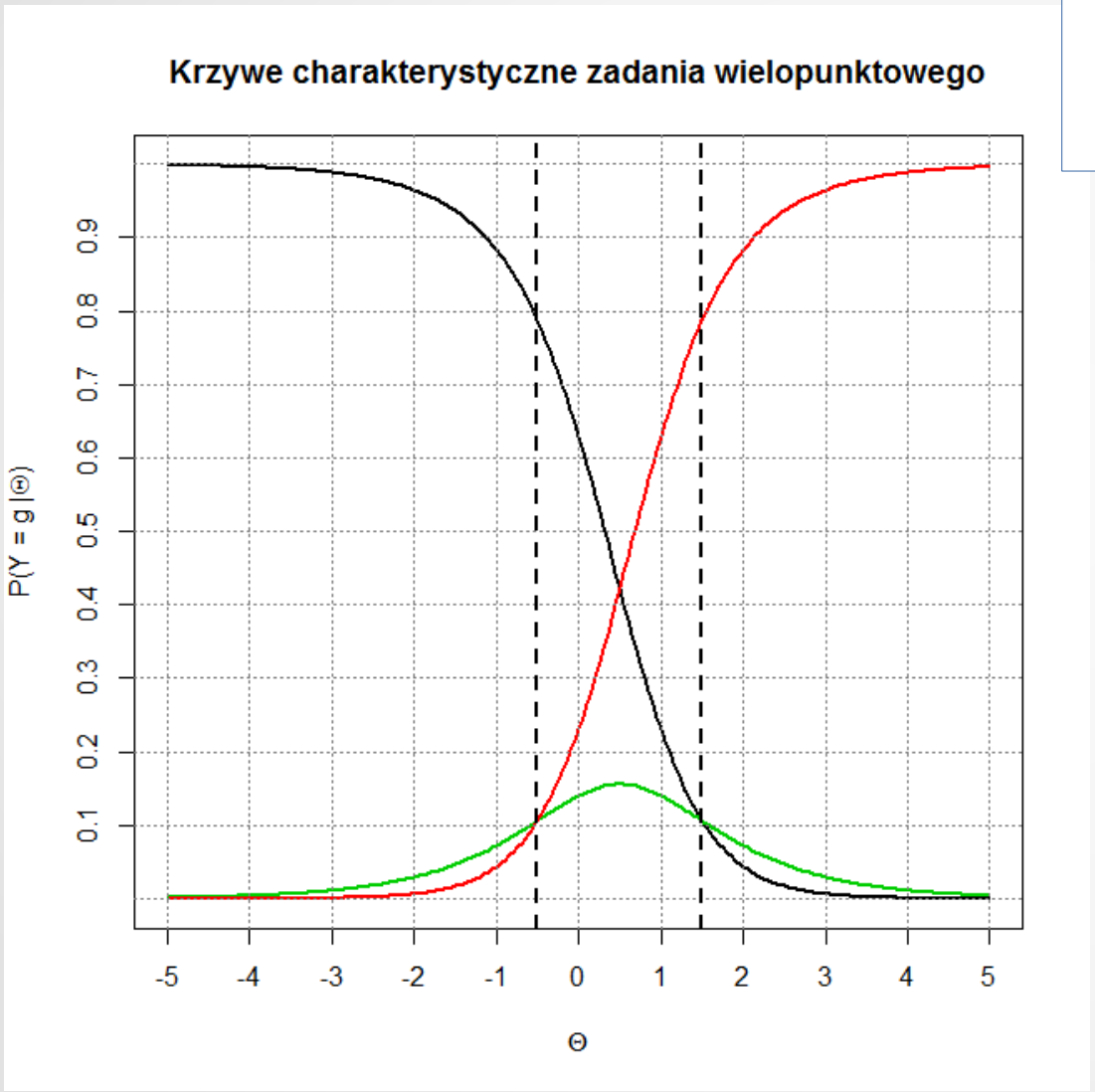


- Za odpowiedź na pytanie prezentowane na wykresie można uzyskać od 0 do 2 punktów.
- Wartości parametrów b_{ik} wynoszą:

  $b_{i1} = -2$; $b_{i2} = 2$;
- Prawdopodobieństwo uzyskania przez ucznia o poziomie umiejętności Θ = 1 liczby punktów:
  - 0 jest równe 0,05;
  - 1 jest równe 0,27;
  - 2 jest równe 0,68.
- Suma: 0,05 + 0,27 + 0,68 = 1

Model Partial Credit

$$P(Y_i = g|\Theta_j) = \frac{\exp\left[\sum_{k=1}^{g} a_i(\Theta - b_{ik})\right]}{1 + \sum_{k=1}^{m_i} P(X_i = k)} \quad \text{dla } g>0$$



Zaburzenie kolejności podpunktów ze względu na trudność:

- Wartości parametrów b_{ik} wynoszą:

  b_{i1} = 1,5; b_{i2} = -0,5;

- W takim przypadku uzyskanie 1 punktu nigdy (dla żadnej wartości mierzonej cechy) nie jest najbardziej prawdopodobne.
- Sytuacja niepożądana – skala oceny zadania nie jest efektywnie wykorzystywana.

Model Graded Response

- W tym modelu nie modelujemy bezpośrednio prawdopodobieństwa uzyskania za zadanie danej liczby punktów, ale nie więcej, niż danej liczby punktów.

$$P\left(X_i \geq 0 | \theta\right) = 1 \qquad P\left(X_i \geq (m+1) | \theta\right) = 0$$

$$P\left(X_i \geq g | \theta\right) = \frac{\exp\left[a_i\left(\theta - b_{ig}\right)\right]}{1+\exp\left[a_i\left(\theta - b_{ig}\right)\right]} \text{ dla } g \in \{1, \ldots, m\}$$

$$P\left(X_i = g | \theta\right) = P\left(X_i \geq g | \theta\right) - P\left(X_i \geq (g+1) | \theta\right)$$

- **Inna niż w modelu Partial Credit interpretacja parametrów b_{ig}:** wskazują na takie wartości mierzonej cechy, dla której prawdopodo-bieństwo uzyskania za zadanie co najmniej *g* punktów wynosi 0,5.
- W odróżnieniu od modeli Partial Credit nie dopuszcza zaburzenia trudności podpunktów ze względu na trudność.
- Za to jest ściśle formalnie powiązany z analizą czynnikową dla zmiennych kategorialnych, estymowanej na podstawie macierzy korelacji polichorycznych.

Modelowanie związku badanej cechy z wynikami zadania o wielu poziomach wykonania

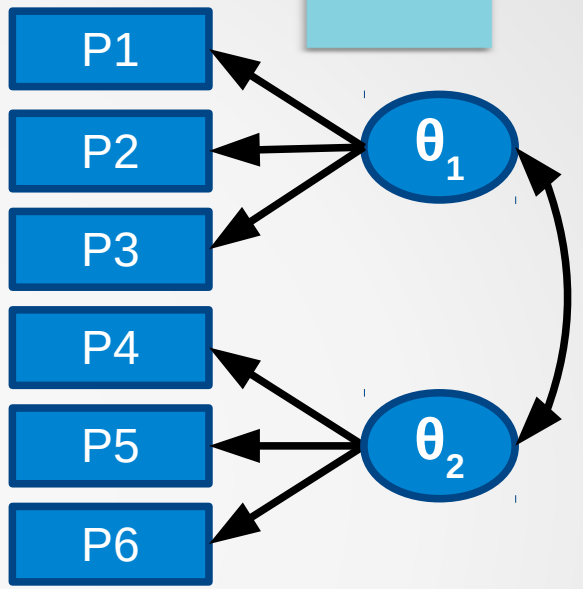
Po co modele wielowymiarowe?

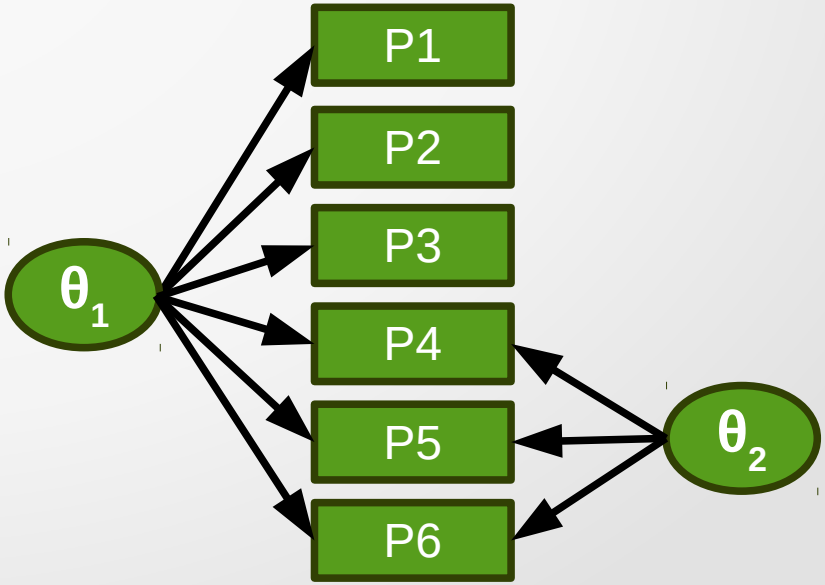
Zwykle jedno z dwóch:

1) Chcemy uzyskać *latentne* oszacowania siły związku pomiędzy cechami (z których każdą mierzymy innym zestawem przejawów).

2) Chcemy modelować złożone uwarunkowania wyników zadań.

   - W celach diagnostycznych (np. wykrywanie naruszenia założenia o jednowymiarowości).
   - W celu uzyskania adekwatnych oszacowań wartości cechy, którą chcemy mierzyć, jeśli w sytuacji badawczej na odpowiedzi wpływały również inne czynniki.





Modele wielowymiarowe - parametryzacja

$$P\left(X_i=1|\theta\right)=1-\frac{1-c_i}{1+\exp\left[\left(\sum_{j=1}^{r} a_{ij}\theta_j\right)+d_i\right]}$$

Aby opisać formalnie zależności pomiędzy kilkoma różnymi cechami a wynikiem zadania potrzeba:

- Wprowadzić dodatkowe współczynniki *dyskryminacji*, które opisują siłę związku danej cechy z wynikiem danego zadania:

  a_{ij} - dyskryminacja *i*-tego zadania ze względu na *j*-tą cechę;

- Przeformułować sposób definiowania *trudności* (która w odróżnieniu od *dyskryminacji* jest cechą zadania, a nie pary cecha-zadanie) – parametr d_i.

  W przypadku jednowymiarowym: $b_i=-\frac{d_i}{a_{i1}}$

Modele wielowymiarowe - parametryzacja

Analogicznie w modelu GPCM:

$$P\left(X_i=g|\theta\right)=\frac{\exp\left[g\left(\sum_{j=1}^{g}a_{ij}\theta_j\right)+d_{ig}\right]}{\sum_{k=0}^{m_i}P\left(X_i=k|\theta\right)}$$

W przypadku jednowymiarowym (dla $g > 0$ i przyjmując $d_{i0} = 0$):

$$b_{ig}=\frac{d_{ig-1}-d_{ig}}{a_{i1}}$$

Dziękujemy za uwagę!

Tomasz Żółtak t.zoltak@ibe.edu.pl
Karolin Świst k.swist@ibe.edu.pl