



EDUKACYJNA WARTOŚĆ DODANA

Jak powstają maturalne wskaźniki EWD?

Tomasz Żółtak

Wiosenna Szkoła EWD 2013

Warszawa, 19-21 kwietnia 2013



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



EDUKACYJNA
WARTOŚĆ
DODANA

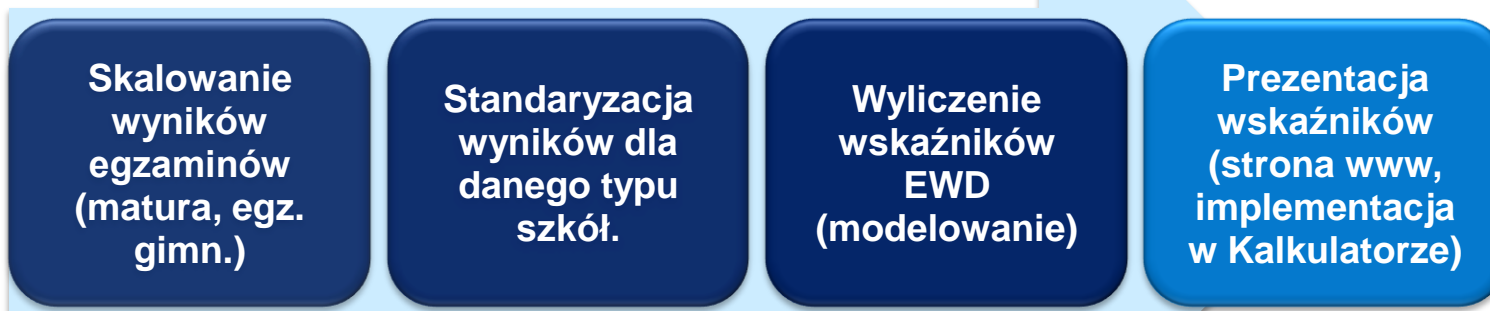
UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY





Plan prezentacji

- Wskaźniki trzyletnie dla szkół kończących się maturą:
 - Skalowanie wyników egzaminów - wskaźniki wyników egzaminacyjnych.
 - Standaryzacja wyników dla danego typu szkół - prezentacja wskaźników wyników egzaminacyjnych na skali 100;15.
 - Modele wykorzystywane do wyliczania trzyletnich wskaźników EWD dla szkół kończących się maturą.
- Jednoroczny wskaźnik matematyczny – Kalkulator maturalny
 - Normalizacja wyników dla danego typu szkół - prezentacja wskaźników wyników egzaminacyjnych na skali 100;15.
 - Modele wykorzystywane do wyliczania jednorocznych wskaźników EWD dla szkół kończących się maturą.





Wskaźniki trzyletnie



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



EDUKACYJNA
WARTOŚĆ
DODANA

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY





EWD a egzamin maturalny – podstawowe problemy

- Różne typy szkół prowadzące do tego samego egzaminu.
- Rozdrobnienie - duża liczba i różnorodność egzaminów.
- Wybór egzaminów zdawanych na maturze.



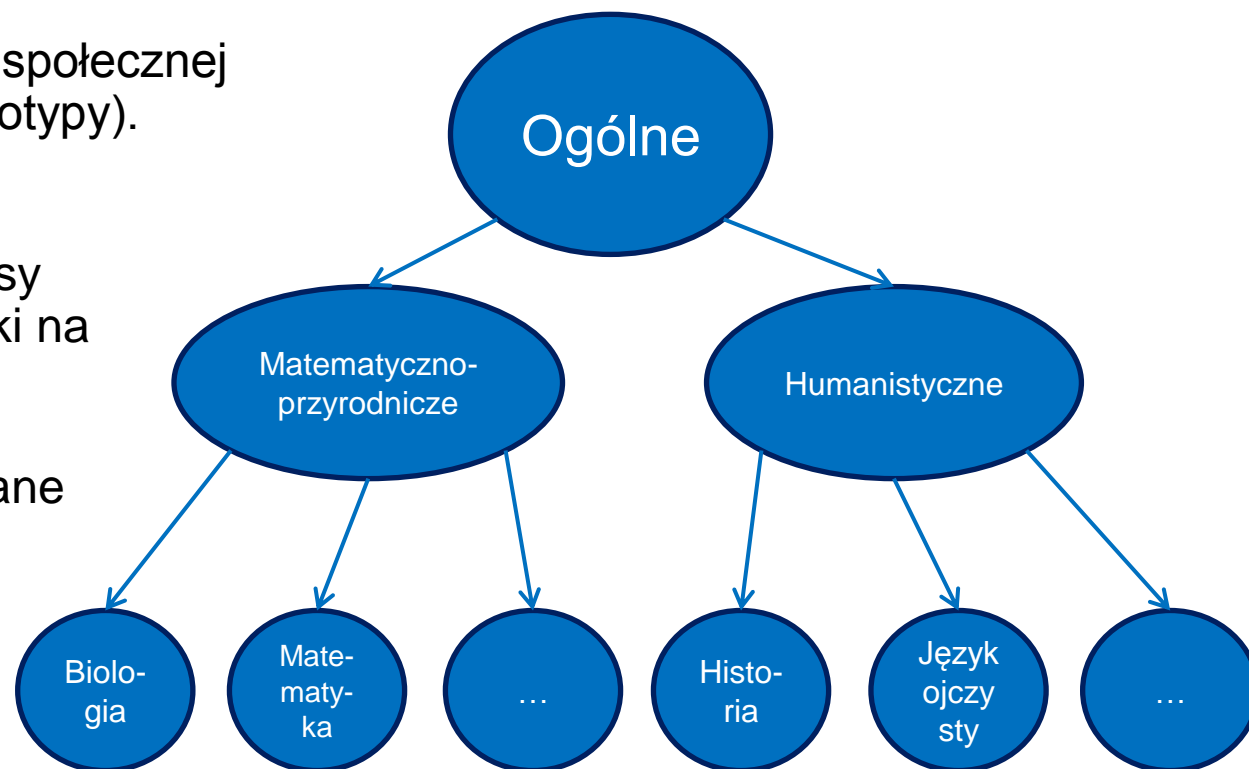
EWD a egzamin maturalny – podstawowe problemy

- Różne typy szkół prowadzące do tego samego egzaminu.
 - Oddzielne wskaźniki EWD dla liceów ogólnokształcących i techników (brak wskaźników dla liceów profilowanych – mała liczba uczniów, zanik tego typu szkół).
- Rozdrobnienie - duża liczba i różnorodność egzaminów.
 - Wykorzystanie złożonych wskaźników wyników egzaminacyjnych, wyliczanych z użyciem metodologii IRT.
- Wybór egzaminów zdawanych na maturze.
 - Uwzględnienie przy wyliczaniu wskaźników wyników egzaminacyjnych parametrów selekcji.



Założenia dotyczące modelowania wskaźnika złożonego

- Zakładamy, że istnieją nieobserwowalne zmienne („cechy ukryte”) odpowiadające za biegłość w danej dziedzinie wiedzy.
- Musimy założyć, że istnieje hierarchiczna struktura umiejętności.
- Założenie to ma odzwierciedlenie w społecznej świadomości (stereotypy).
- Jest potwierdzone instytucjonalne (klasy profilowane, kierunki na uczelniach).
- Jest udokumentowane w badaniach.





Wykorzystywane wskaźniki wyników egzaminacyjnych

■ Wskaźniki wyników matury:

- W zakresie **języka polskiego** (poziom podst. i rozsz.).
- W zakresie **matematyki** (poziom podst. i rozsz.).
- W zakresie **przedmiotów humanistycznych** (wszystkie przedmioty na poziomie podst. i rozsz.):
 - (j. polski,
 - historia,
 - WOS.
- W zakresie **przedmiotów matematyczno-przyrodniczych** (wszystkie przedmioty na poziomie podst. i rozsz.):
 - biologia,
 - chemia,
 - fizyka,
 - geografia,
 - informatyka,
 - matematyka.



Problem konstrukcji wskaźnika złożonego

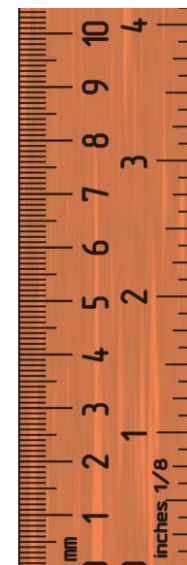
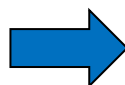
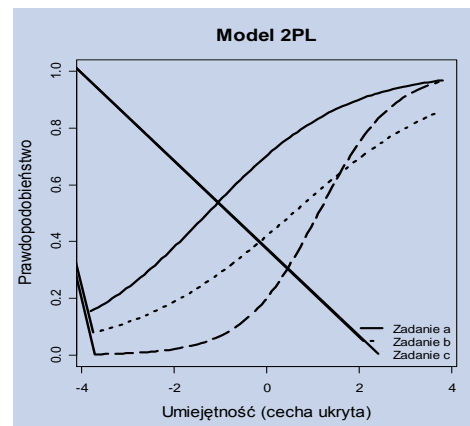
	matematyka					biologia					Informatyka					
uczeń	Zad_1	Zad_2	Zad_3	Zad_4	(...)	Zad_1	Zad_2	Zad_3	Zad_4	(...)	Zad_1	Zad_2	Zad_3	Zad_4	(...)	Wynik ogólny
001	1	1	0	0		1	1	1	1		-	-	-	-		?
002	1	1	1	1		1	1	1	1		1	0	1	1		?
003	1	1	0	0		1	1	1	0		0	0	1	1		?
004	1	0	1	1		1	1	0	0		0	1	0	0		?
005	0	0	1	1		-	-	-	-		-	-	-	-		?
006	1	1	0	0		-	-	-	-		0	1	0	0		?
007	1	1	1	1		-	-	-	-		-	-	-	-		?
008	1	1	1	0		-	-	-	-		-	-	-	-		?



Model IRT 2PL/GPCM proces skalowania

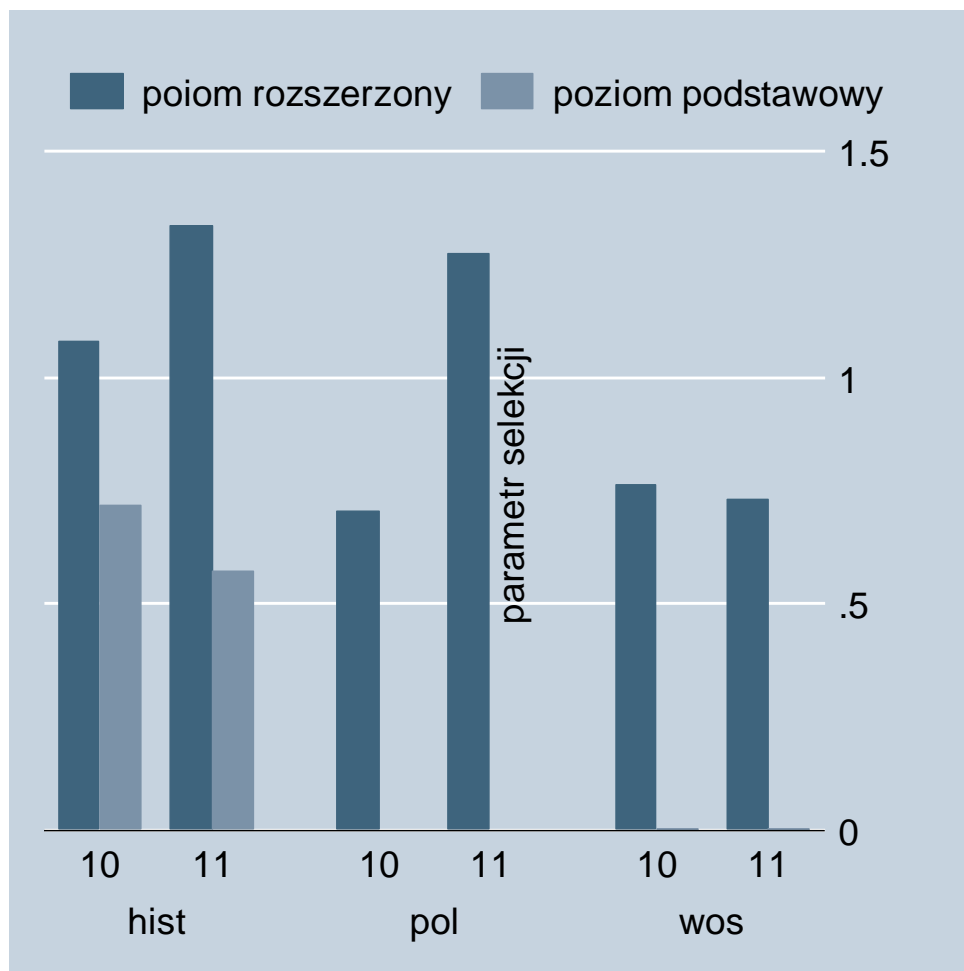
$$P_{ik}(\theta_n) = \frac{\exp(a_i(\theta_n - (\beta_i + \delta_k)))}{1 + \exp(a_i(\theta_n - (\beta_{ik} + \delta_k)))}$$

Id	Zad 1	Zad 2	Zad 3	Zad 4	Zad 5	Zad 6	Zad 7	Zad 8	Zad 9
001	1	1	0	0	1	1	1	1	1
002	1	0	0	1	0	1	2	3	1
003	1	1	1	1	0	0	1	3	1
004	0	0	1	1	0	1	0	1	1
005	0	0	1	1	0	0	2	0	0
006	1	1	0	0	1	2	2	4	2
007	1	1	1	1	1	1	2	4	1
008	1	1	1	1	1	1	2	3	1





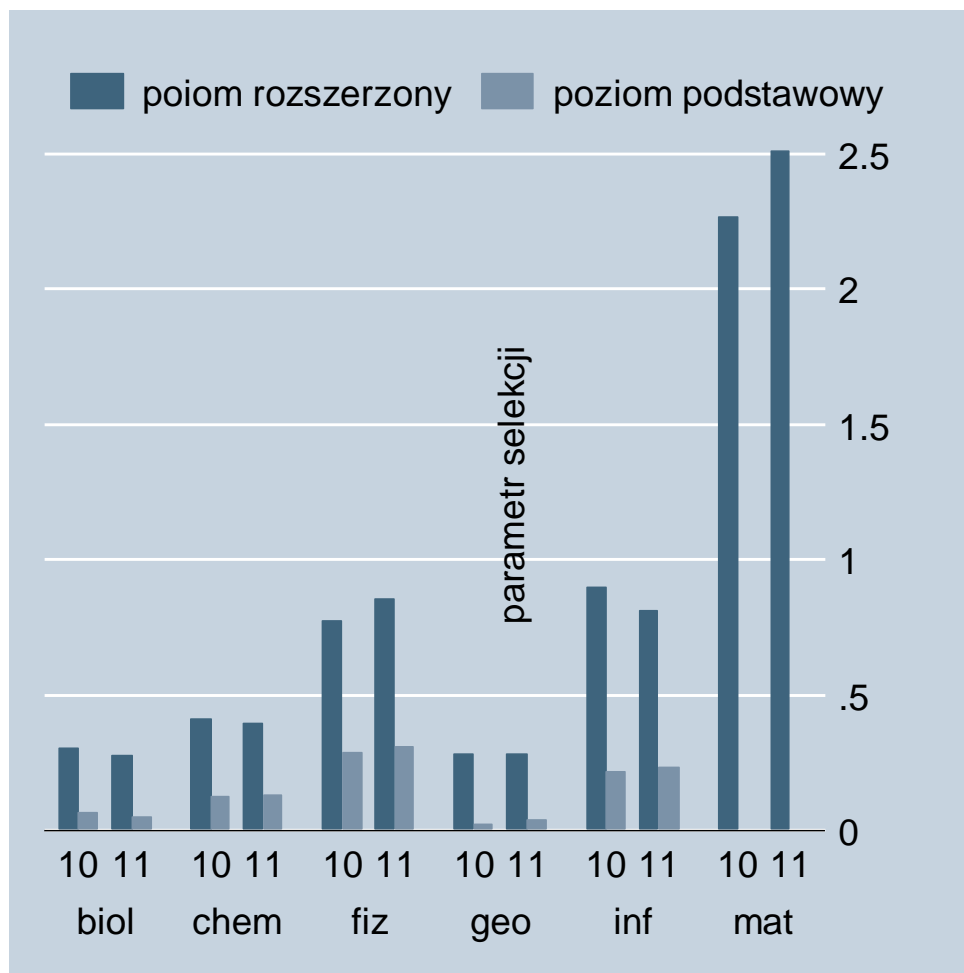
Parametry selekcji: przedmioty humanistyczne



- Parametry selekcji pokazują, jak silnie wybór zdawania danego przedmiotu na danym poziomie jest skorelowany z poziomem umiejętności uczniów – im większa wartość parametru tym zależność jest silniejsza (do arkusza częściej podchodzą uczniowi mocni a rzadziej słabi).
- Pozwala to przy przewidywaniu poziomu umiejętności uczniów wziąć poprawkę na to, że arkusz, do którego nie podchodzili słabi uczniowie wydaje się łatwiejszy, a arkusz do którego nie podchodzili uczniowie mocniejsi, wydaje się trudniejszy.
- Arkusze na poziomie rozszerzonym mają wyższe wartości parametrów selekcji, niż arkusze na poziomie podstawowym.



Parametry selekcji: przedmioty matematyczno-przyrodnicze



- Parametry selekcji pokazują, jak silnie wybór zdawania danego przedmiotu na danym poziomie jest skorelowany z poziomem umiejętności uczniów – im większa wartość parametru tym zależność jest silniejsza (do arkusza częściej podchodzą uczniowi mocni a rzadziej słabi).
- Pozwala to przy przewidywaniu poziomu umiejętności uczniów wziąć poprawkę na to, że arkusz, do którego nie podchodzili słabi uczniowie wydaje się łatwiejszy, a arkusz do którego nie podchodzili uczniowie mocniejsi, wydaje się trudniejszy.
- Arkusze na poziomie rozszerzonym mają wyższe wartości parametrów selekcji, niż arkusze na poziomie podstawowym.



Co jest brane pod uwagę przy wyliczaniu wyskalowanego wyniku ucznia?

■ Własności pomiarowe zadań:

- W procesie skalowania dla każdego zadania szacowane są dwa parametry: trudność i dyskryminacja. Drugi z nich zdaje sprawę z tego, na ile dobrze zadanie pozwala różnicować uczniów ze względu na poziom umiejętności mierzonej danym testem.
- Przy wyliczaniu wyskalowanego wyniku ucznia **przykładamy większą wagę do zadań o wysokiej dyskryminacji**. Uznajemy, że dobre rozwiązanie zadania o dobrych własnościach pomiarowych jest lepszym świadectwem wysokiego poziomu umiejętności.
- W przypadku **języka polskiego** skutkuje to – w praktyce - **przykładaniem niższej wagi do wyników wypracowania**, gdyż oceny wypracowania co do zasady okazują się mieć wyraźnie niższą dyskryminację niż zadania testowe z języka polskiego. Może to skutkować wyraźnymi różnicami w stosunku do surowych wyników – jeśli uczeń napisał wypracowanie wyraźnie gorzej lub test, albo odwrotnie.
- W przypadku innych przedmiotów brak tego rodzaju systematycznych zależności.

■ Trudność arkuszy, które rozwiązywał uczeń:

- Trudność arkuszy ustalana jest przy uwzględnieniu, że niektóre z nich rozwiązywali – ogólnie rzecz biorąc – *mocniejsi*, a inne *słabsi* uczniowie.
- Jeśli uczeń rozwiązuje trudniejszy arkusz *oczekujemy* od niego niższego wyniku.
- Uczniowie nie są *karani* za podejście do trudniejszego przedmiotu (jeśli piszą go na miarę swoich możliwości).

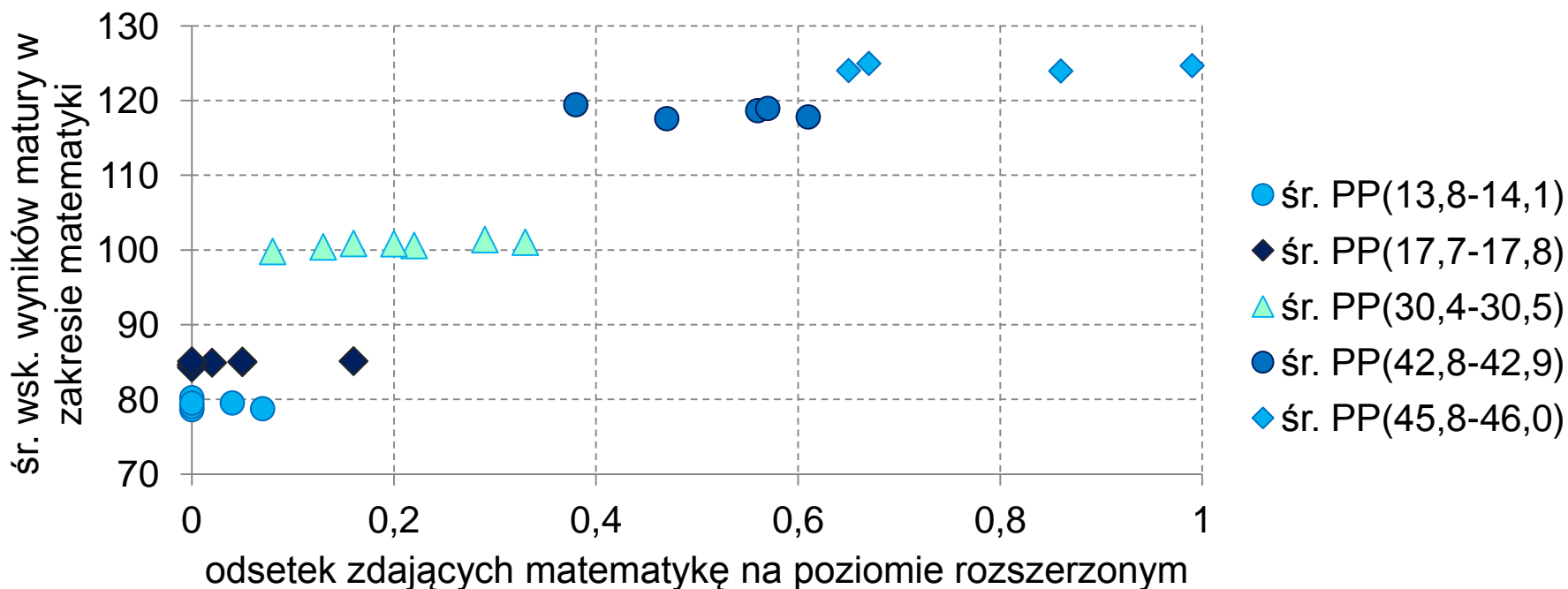


Czy duża liczba zdających PR zaniża wyniki?

Przykład empiryczny - matura z matematyki z 2011 r.

- poziom podstawowy obowiązkowy dla wszystkich – na podstawie jego wyników można ocenić średni poziom umiejętności wszystkich zdających maturę w danej szkole;
- wybrano pięć grup szkół – w ramach każdej grupy szkoły są niemal identyczne ze względu na średnie wyniki (surowe) matury z matematyki na poziomie podstawowym (różnice $<0,3$ pkt.);

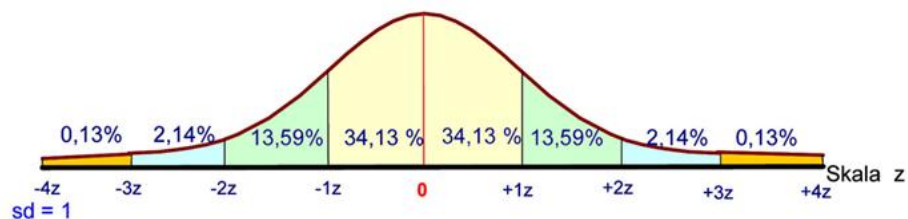
W ramach grup (a więc przy kontroli wyników z poziomu podstawowego) odsetek uczniów wybierających zdawanie egzaminu na poziomie rozszerzonym nie ma znaczenia dla średniej wyskalowanych wyników szkoły.



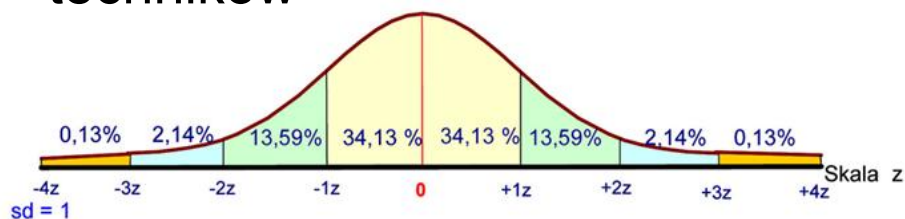


Standaryzacja

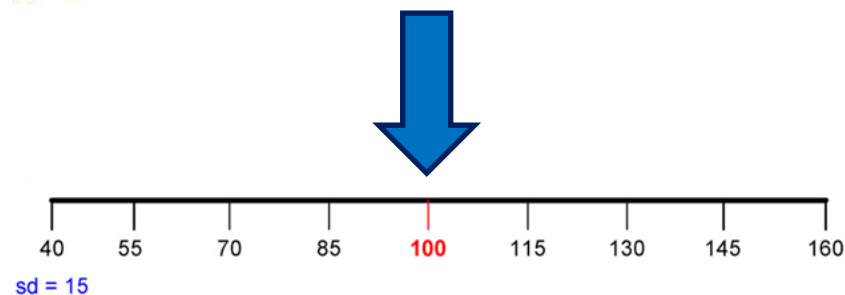
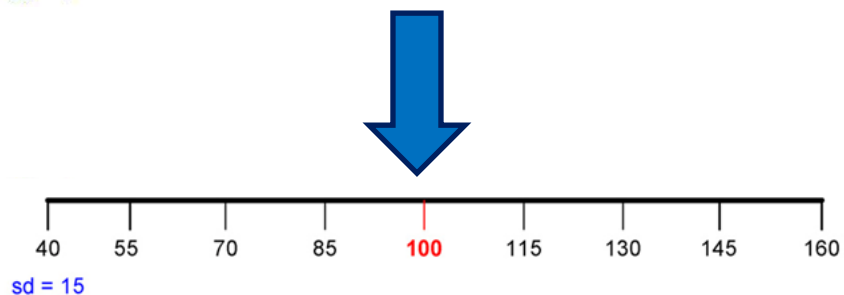
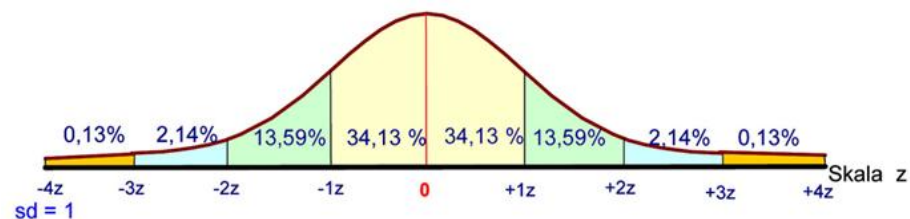
wyniki skalowania – wszyscy zdający



uczniowie
techników



uczniowie liceów
ogólnokształcących





Wskaźniki wyników egzaminu gimnazjalnego

- W zakresie części humanistycznej.
 - W zakresie części matematyczno-przyrodniczej.
- (Dopiero w 2015 r. LO ukończą osoby, które zdawały egz. gimn. w nowej formule.)
- Skalowane IRT modelem 2PL/GPCM, jak wskaźniki wyników matury.
 - Nie ma tu problemy z wybieralnością przedmiotów – bez wskaźników selekcji.

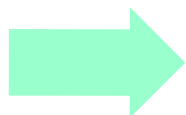


Skąd się biorą wskaźniki trzyletnie?

Licea ogólnokształcące:

matura 2010

- egz. gimn. 2007
- egz. gimn. 2006
(wydłu. tok kształcenia)



matura 2011

- egz. gimn. 2008
- egz. gimn. 2007
(wydł. tok kształcenia)



matura 2012

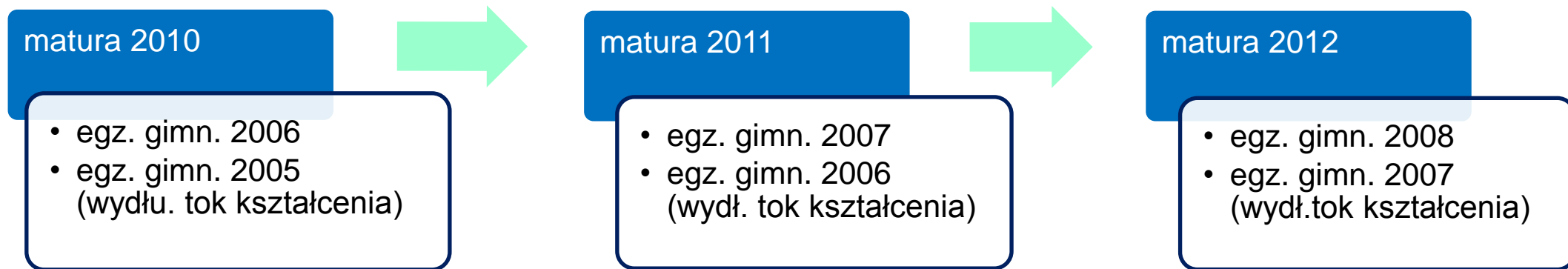
- egz. gimn. 2009
- egz. gimn. 2008
(wydł.tok kształcenia)

- Trzy kolejne sesje egzaminacyjne.
- Pięć lat pracy liceum.
 - W odniesieniu do uczniów o toku kształcenia wydłużonym o rok sześć lat pracy.



Skąd się biorą wskaźniki trzyletnie?

Technika:



- Trzy kolejne sesje egzaminacyjne.
- Sześć lat pracy technikum.
 - W odniesieniu do uczniów o toku kształcenia wydłużonym o rok siedem lat pracy.



Trzyletnie maturalne wskaźniki EWD

EG część
humanistyczna
(skalowany IRT)

5(+1) lat pracy LO,
6(+1) lat pacy
technikum

**Maturalny
wskaźnik
z j. polskiego**
(j. pol: pp i pr)

EG część
humanistyczna
(skalowany IRT)

5(+1) lat pracy LO,
6(+1) lat pacy
technikum

**Maturalny
wskaźnik
humanistyczny**
(hist, j. pol, WOS:
pp i pr)

EG część mat.-
przyrod.
(skalowany IRT)

5(+1) lat pracy LO
6(+1) lat pacy
technikum

**Maturalny
wskaźnik
matematyczny**
(mat: pp i pr)

EG część mat.-
przyrod.
(skalowany IRT)

5(+1) lat pracy LO,
6(+1) lat pacy
technikum

**Maturalny
wskaźnik mat.-
przyr.**
(biol, geo, fiz, chem
mat, inf: pp i pr)



Wykluczenia

Przed przystąpieniem do modelowania usuwane są dane dotyczące:

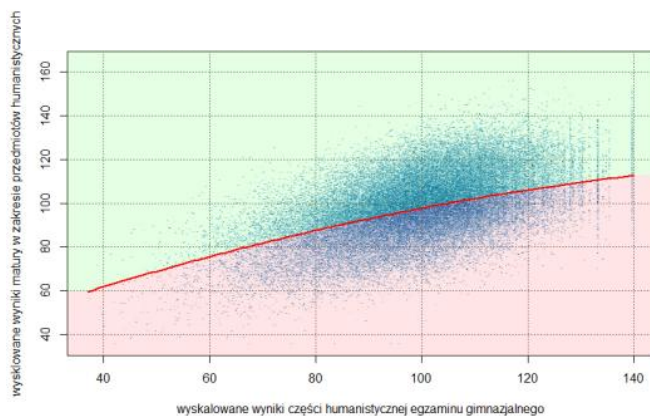
- Uczniów szkół specjalnych i dla dorosłych.
- Uczniów liceów profilowanych.
- Uczniów, o toku kształcenia wydłużonym o więcej niż rok.
- Osoby, które zdając maturę były starsze niż:
 - 22 lata w przypadku liceów,
 - 23 lata w przypadku techników.



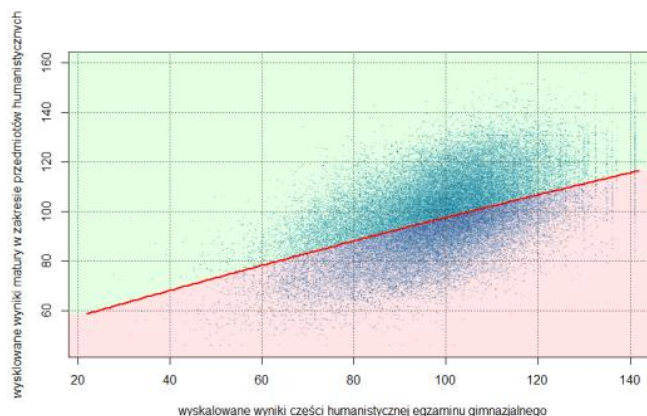
Modele używane do wyliczania wskaźników trzyletnich

Dla każdej z 6 grup uczniów – wyznaczanych przez przecięcie roku zdawania matury i roku zdawania egz. gimn. – wyznacza się oddzielną linię przewidywania.

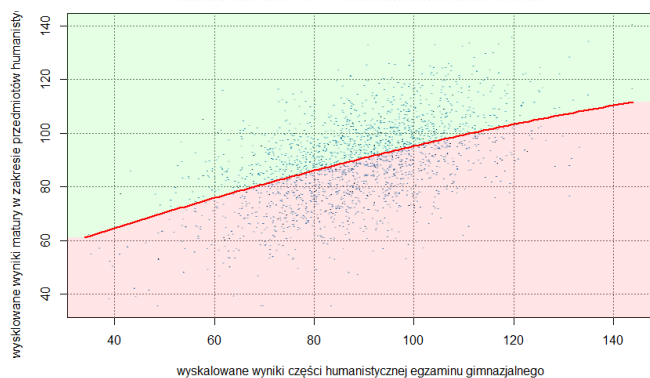
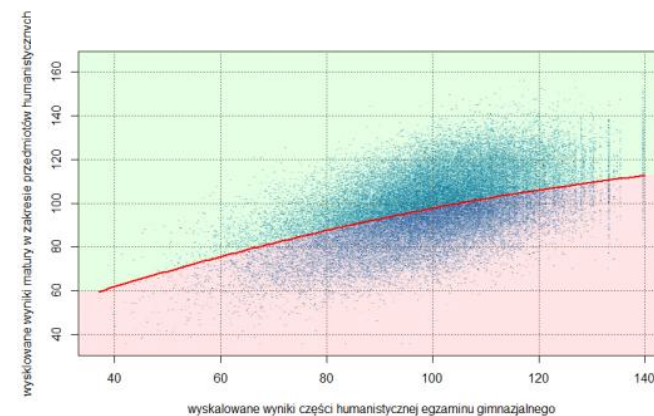
egz. gimn. 2007 → matura 2010



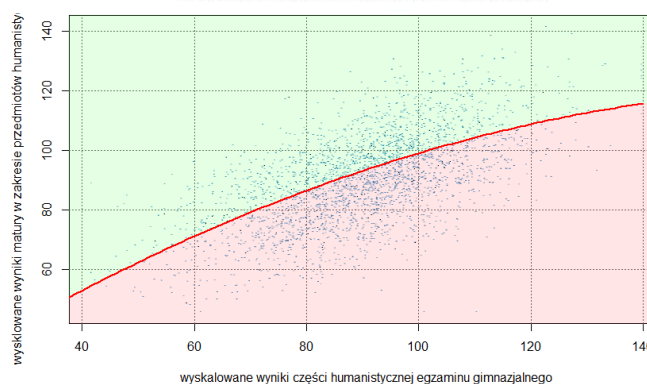
egz. gimn. 2008 → matura 2011



egz. gimn. 2009 → matura 2012

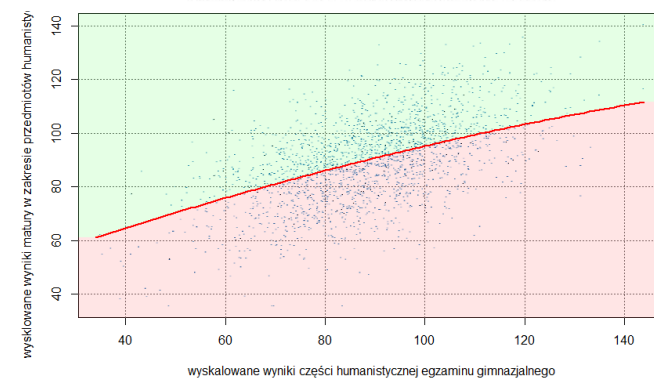


egz. gimn. 2006 → matura 2010



egz. gimn. 2007 → matura 2011

ii



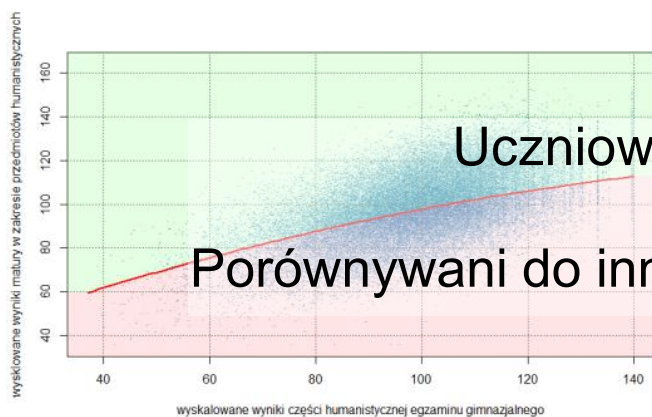
egz. gimn. 2008 → matura 2012



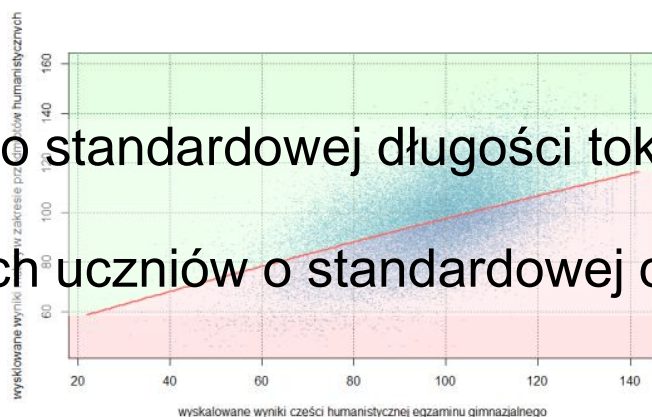
Modele używane do wyliczania wskaźników trzyletnich

Dla każdej z 6 grup uczniów – wyznaczanych przez przecięcie roku zdawania matury i roku zdawania egz. gimn. – wyznacza się oddzielną linię przewidywania.

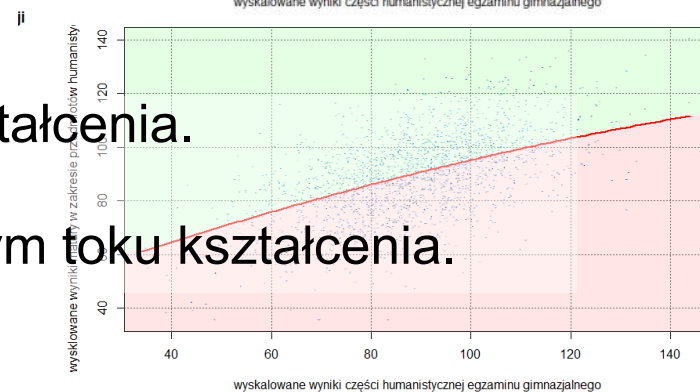
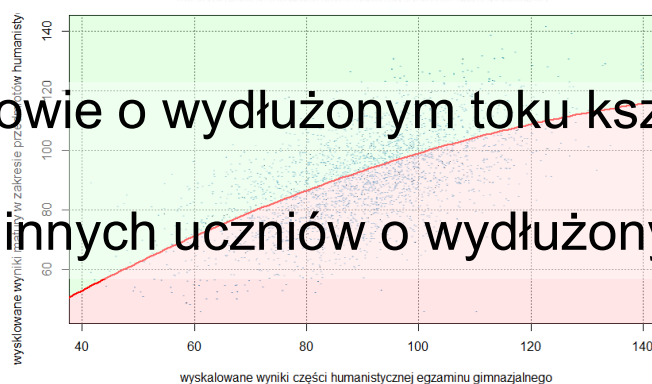
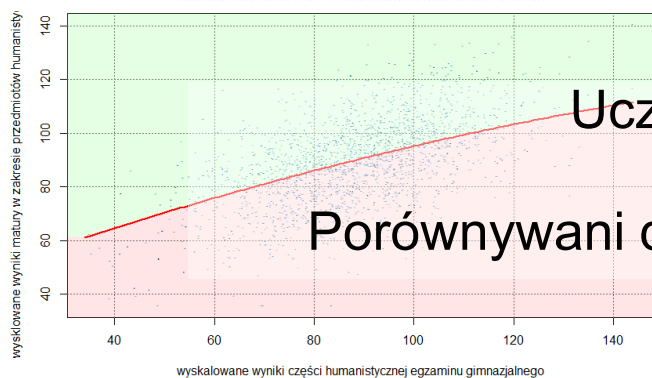
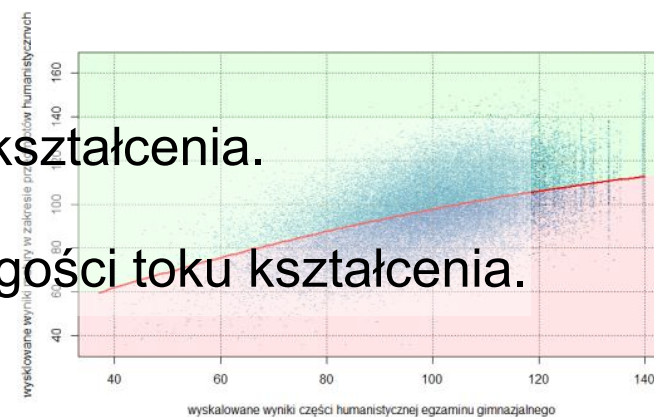
egz. gimn. 2007 → matura 2010



egz. gimn. 2008 → matura 2011



egz. gimn. 2009 → matura 2012



egz. gimn. 2006 → matura 2010

egz. gimn. 2007 → matura 2011

egz. gimn. 2008 → matura 2012



Modele używane do wyliczania wskaźników trzyletnich

- Oprócz wyników (odpowiedniej części) egzaminu gimnazjalnego do przewidywania wyników (danej części) matury używane są informacje o:
 - płci,
 - posiadaniu zaświadczenia o dysleksji na egzaminie gimnazjalnym i/lub maturze.
- Stosowany model regresji określany jest mianem regresji mieszanych efektów lub regresją wielopoziomową.
- EWD szkoły wyliczane jest na podstawie reszt regresji (różnic pomiędzy wynikami, jakie uczniowie osiągnęli na maturze a wynikami przewidywanymi na podstawie modelu):
 - EWD jest średnią reszt regresji w ramach szkoły *ściągniętą* w kierunku zera – w tym większym stopniu, im szkoła jest mniejsza.



Wskaźniki trzyletnie – uwagi interpretacyjne

- Dla interpretacji wskaźników humanistycznego i matematyczno-przyrodniczego ważna jest wiedza o liczbie uczniów, którzy zdawali poszczególne przedmioty na poszczególnych poziomach.
- Wskaźniki EWD opisują szkołę w zakresie nauczania przedmiotów, które jej uczniowie zdają na maturze.
- Największy wpływ na wartości wskaźników mają te przedmioty, które zdaje najwięcej uczniów danej szkoły.
- Informacje o liczbie zdających poszczególne przedmioty na poziomie podstawowym i rozszerzonym publikowana jest na stronie internetowej razem ze wskaźnikami EWD.



Jednoroczny wskaźnik w zakresie matematyki na poziomie podstawowym



Jednoroczny wskaźnik EWD w zakresie matematyki – podstawowe cechy

- Tylko jeden wskaźnik – obejmujący wyłącznie wyniki matury z matematyki na poziomie podstawowym (na wejściu część mat.-przyr. egzaminu gimnazjalnego).
 - Ograniczone zastosowanie do szkół o bardzo wysokich wynikach z matematyki – efekt sufitowy matury z matematyki na poziomie podstawowym.
- Tak samo unormowana skala wyników egzaminów (100;15), ale zupełnie inny niż we wskaźnikach trzyletnich sposób wyliczania wyników:
 - Normalizacja ekwwantylowa zamiast IRT.
 - Możliwość łatwego przeliczenia wyników z sumy *surowych* punktów, lub wyniku podanego na skali centylowej.
- Przeznaczony do analiz wewnątrzszkolnych.
 - Zaimplementowany w *Kalkulatorze*, a nie udostępniany za pośrednictwem strony internetowej.



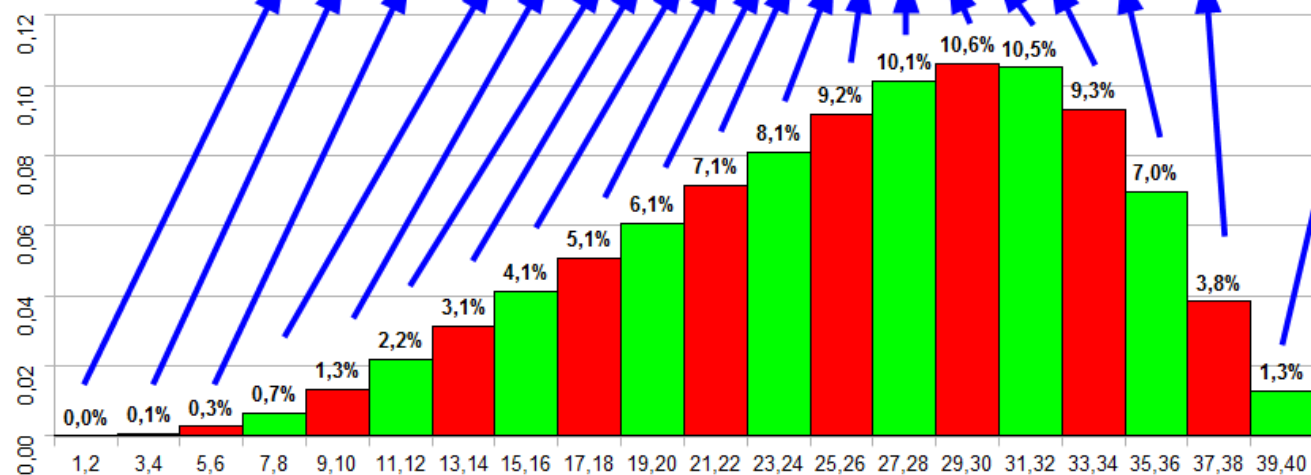
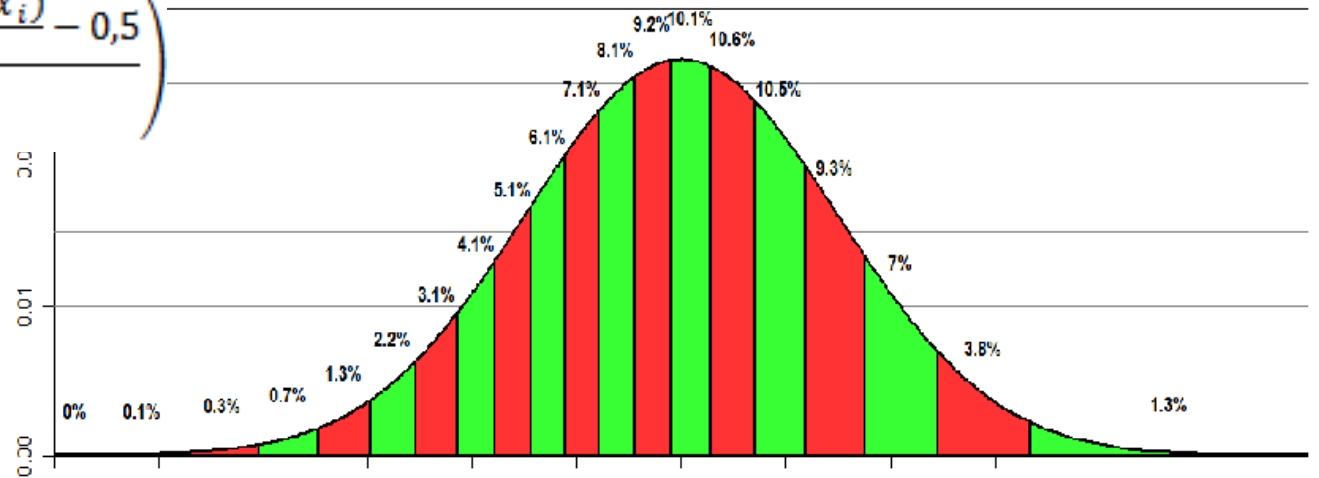
Normalizacja ekwwantylowa

$$U(X = x_i) = 100 + 15 \cdot \Phi^{-1} \left(\frac{N(X \leq x_i) - \frac{N(X = x_i)}{2} - 0,5}{n} \right)$$

$U(X=x_i)$ - wynik znormalizowany dla wyniku surowego równego x_i ;
 Φ^{-1} - funkcja odwrotna do dystrybuanty rozkładu normalnego standaryzowanego;
 $N(X \leq x_i)$ - liczba zdających z wynikiem niższym niż x_i ;
 $N(X = x_i)$ - liczba zdających z wynikiem surowym równym x_i ;
 n - liczba zdających

przedziały
różnej
szerokości

Funkcja gęstości prawdopodobieństwa rozkładu normalnego (100; 15)

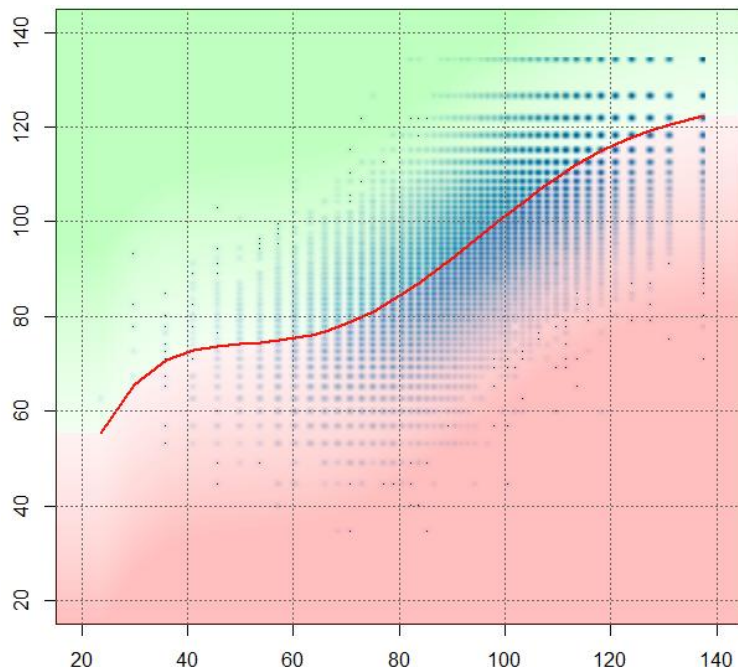


przedziały
równiej
szerokości

Rozkład sprawdzianu 2011 zrekodowanego na skalę 1-20



Modele używane do wyliczania wskaźników trzyletnich



- Oddzielne modele dla LO i T.
- Wykluczenia zdających z modelowania – jak w modelach trzyletnich.
- Regresja metodą najmniejszych kwadratów.
- Oddzielnie wyznaczana linia przewidywania dla uczniów o standardowej długości toku kształcenia i uczniów o toku kształcenia wydłużonym o 1 rok.
- Uwzględnia informacje o płci i posiadaniu zaświadczenia o dysleksji na egzaminie gimnazjalnym i/lub maturze.
- EWD szkoły wyliczane jako średnia reszt regresji (różnic pomiędzy wynikami, jakie uczniowie osiągnęli na maturze a wynikami przewidywanymi na podstawie modelu).



EDUKACYJNA WARTOŚĆ DODANA

Instytut Badań Edukacyjnych

ul. Górczewska 8, 01-180 Warszawa

tel.: (22) 241 71 00, e-mail: ewd@ewd.edu.pl



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



EDUKACYJNA
WARTOŚĆ
DODANA

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

