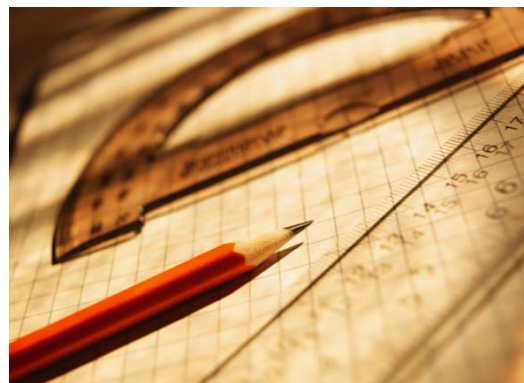




## Interpretacja wyników standaryzowanych testów osiągnięć szkolnych na przykładzie skali 100/15



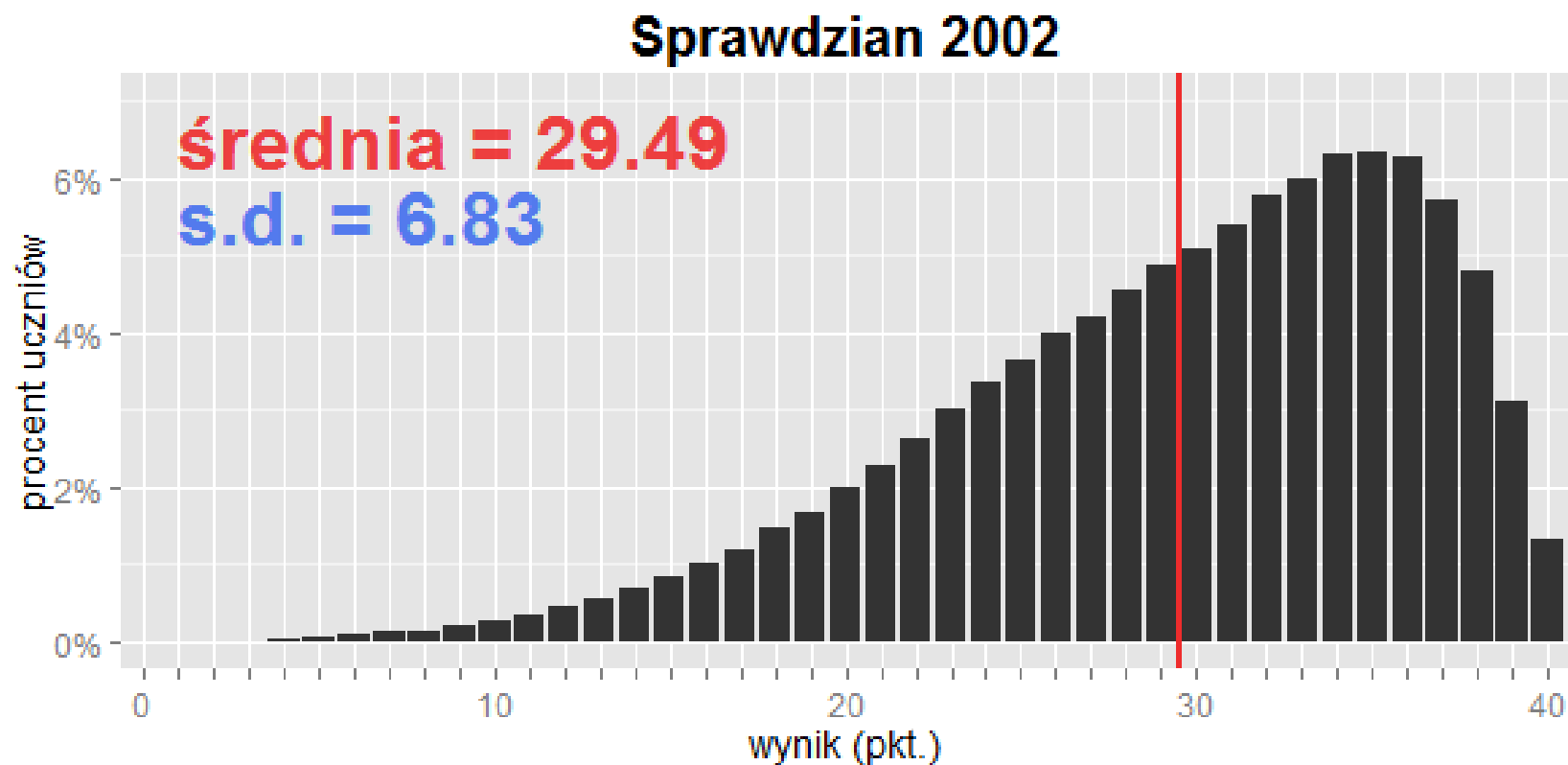
**Michał Modzelewski**  
**Roman Dolata**

Zespół Szkolnych Uwarunkowań Efektywności Kształcenia

Warszawa, 28 listopada 2014 r.



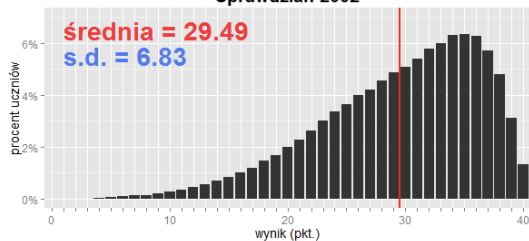
# Wyniki *Sprawdzianu* w klasie szóstej 2002



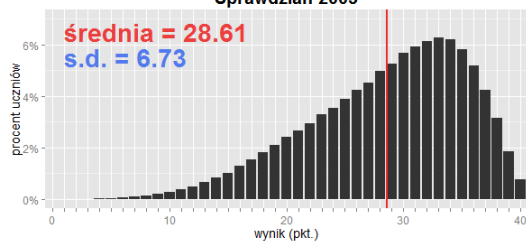


# Wyniki Sprawdzianu w klasie szóstej 2002-2014

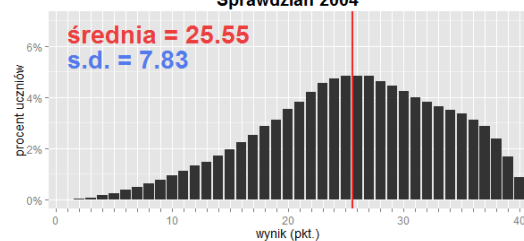
Sprawdzian 2002



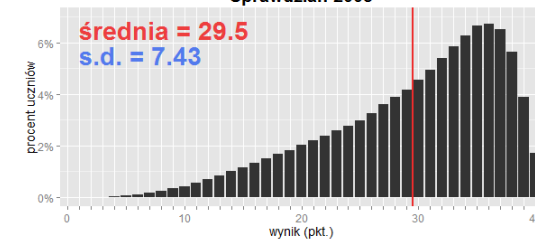
Sprawdzian 2003



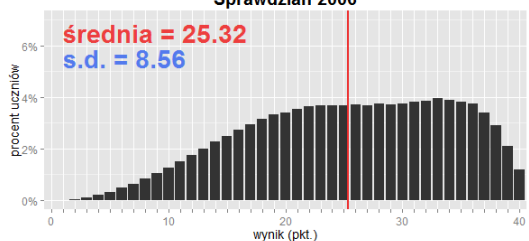
Sprawdzian 2004



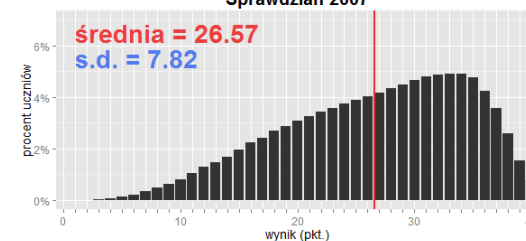
Sprawdzian 2005



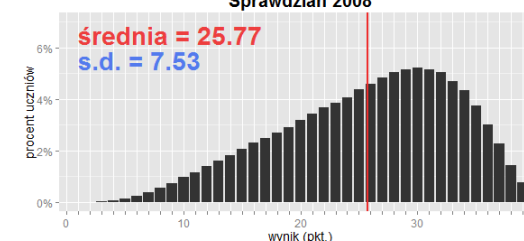
Sprawdzian 2006



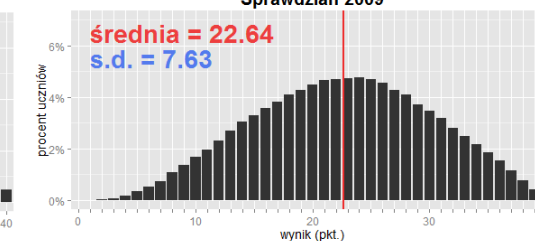
Sprawdzian 2007



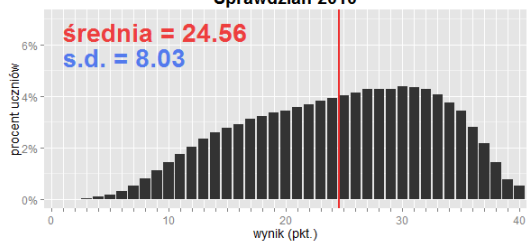
Sprawdzian 2008



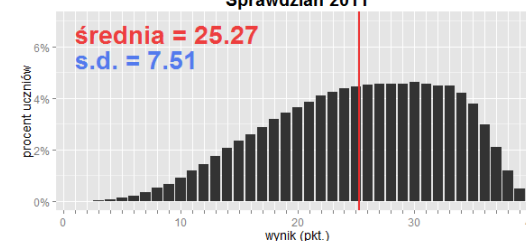
Sprawdzian 2009



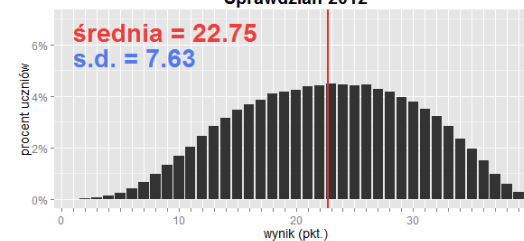
Sprawdzian 2010



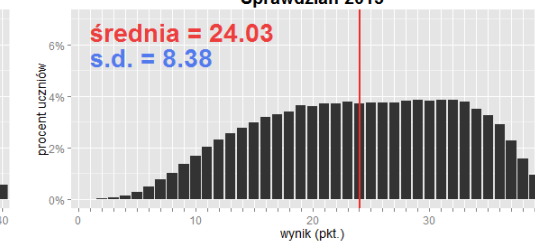
Sprawdzian 2011



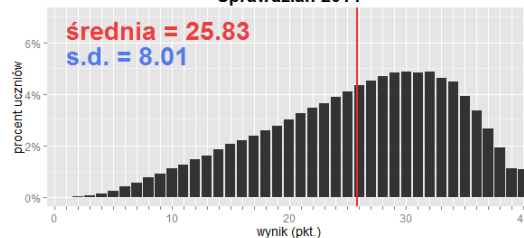
Sprawdzian 2012



Sprawdzian 2013

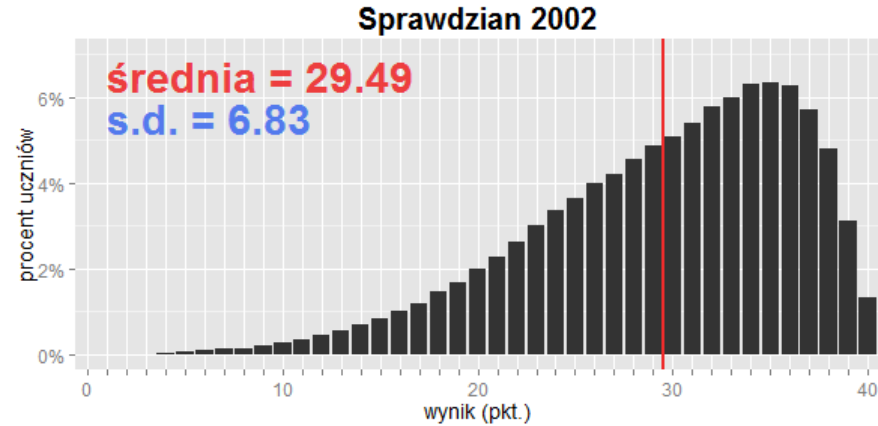


Sprawdzian 2014



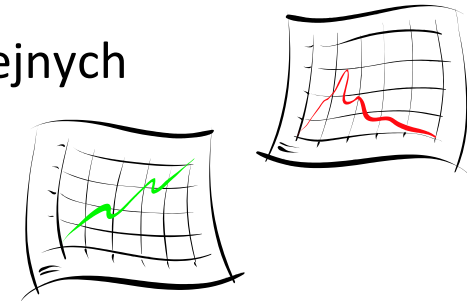


# Skąd się biorą zróżnicowane rozkłady wyników surowych?



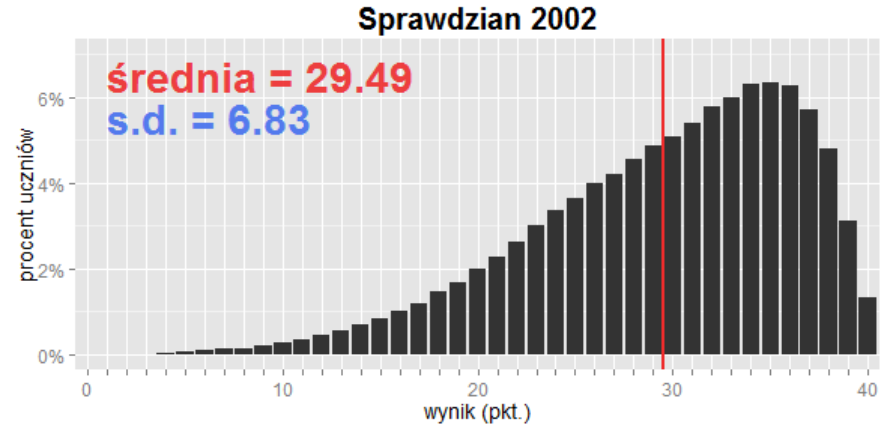
Źródła zmienności rozkładów wyników surowych dla kolejnych roczników:

- zmiany w rozkładzie umiejętności w kolejnych rocznikach absolwentów SP
- <http://pwe.ibe.edu.pl>





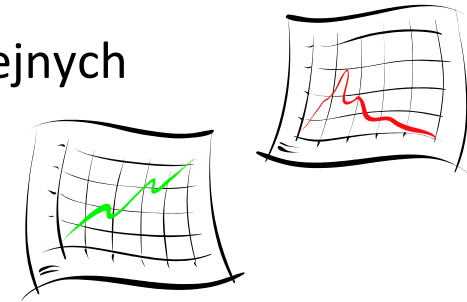
# Skąd się biorą zróżnicowane rozkłady wyników surowych?



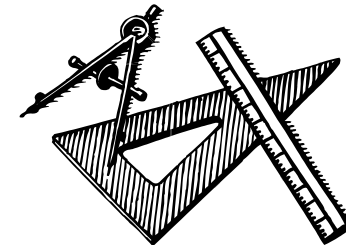
Źródła zmienności rozkładów wyników surowych dla kolejnych roczników:

- zmiany w rozkładzie umiejętności w kolejnych rocznikach absolwentów SP

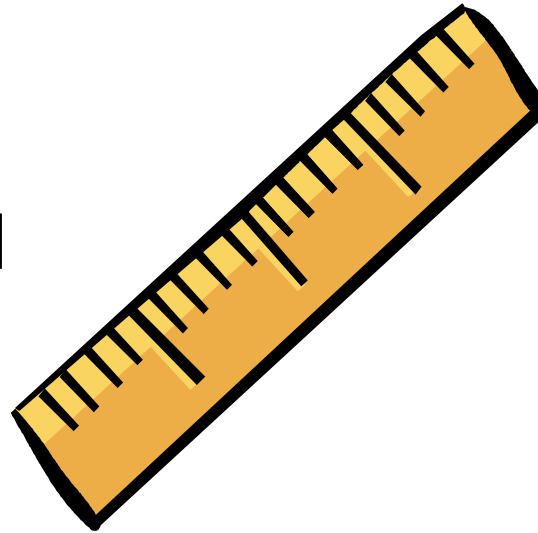
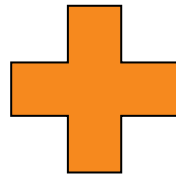
- <http://pwe.ibe.edu.pl>



- zmiany narzędzia pomiarowego, czyli testu



# Na jakiej skali przedstawiać wyniki testów?



1. Kwestia pożądanego kształtu rozkładu wyników
2. Kwestia odpowiedniej skali wyników



# Zakładany kształt rozkładu wyników testu...

... a funkcja testu.

Liczba wyników Rodzaj testu	Niskich	Przeciętnych	Wysokich
Test minimum kompetencji	+	++	+++
Test szerokiego spektrum	++	+++	++
Test selekcyjny	+++	++	+



# Zakładany kształt rozkładu wyników testu...

... a funkcja testu.

Liczba wyników Rodzaj testu	Niskich	Przeciętnych	Wysokich
Test minimum kompetencji	+	++	+++
Test szerokiego spektrum	++	+++	++
Test selekcyjny	+++	++	+

Sprawdzian, egzamin gimnazjalny, testy diagnostyczne typu OBUT  
to testy w założeniu badające szerokie spektrum umiejętności





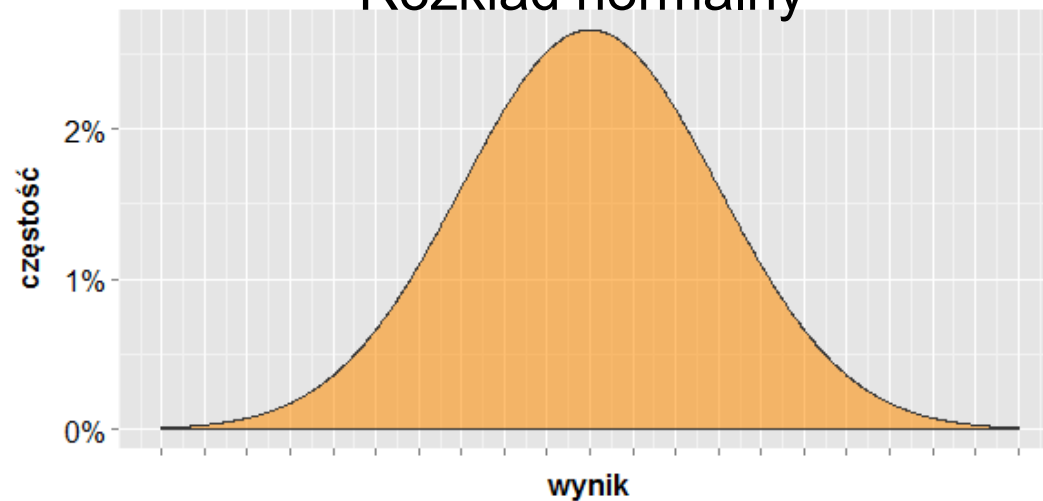
# Zakładany kształt rozkładu wyników testu...

... a funkcja testu.

Liczba wyników \ Rodzaj testu	Niskich	Przeciętnych	Wysokich
Test minimum kompetencji	+	++	+++
Test szerokiego spektrum	++	+++	++
Test selekcyjny	+++	++	+

Sprawdzian, egzamin gimnazjalny, testy diagnostyczne typu OBUT to testy w założeniu badające szerokie spektrum umiejętności

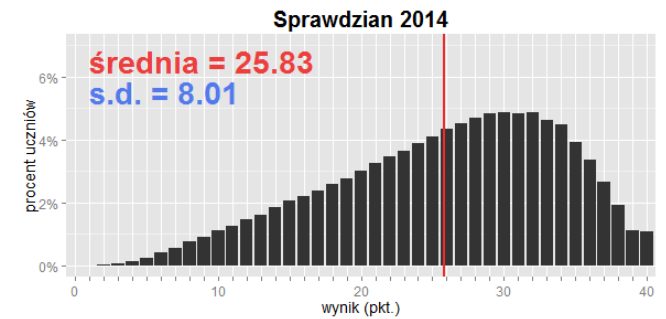
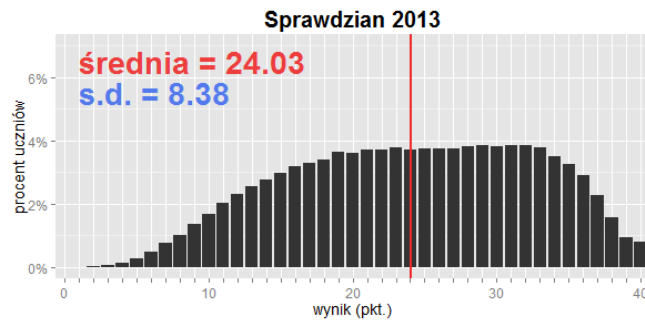
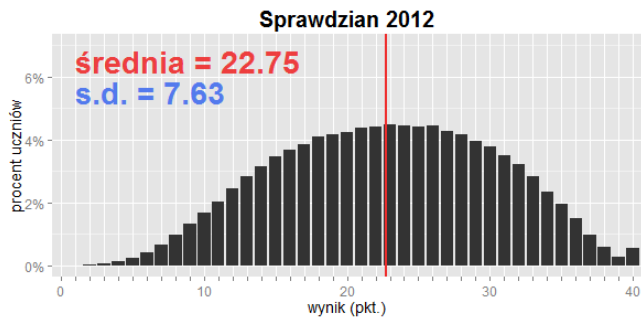
Rozkład normalny





# Zakładany kształt rozkładu wyników testu...

... a faktyczny rozkład wyników.

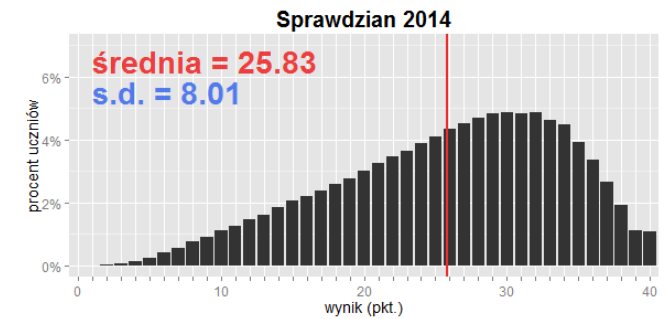
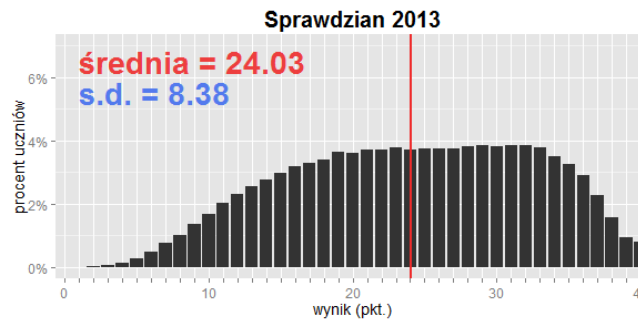
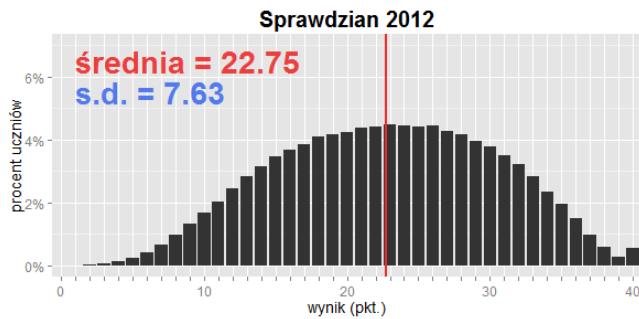


W praktyce bywa różnie 😊



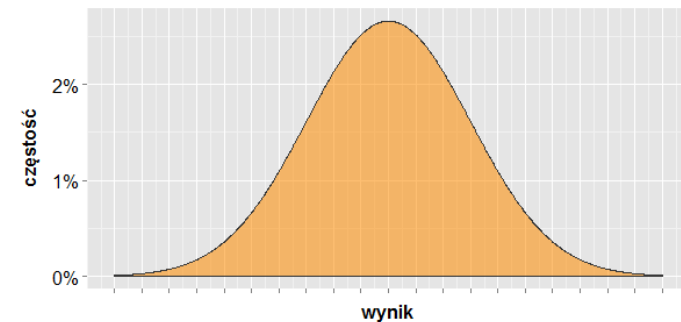
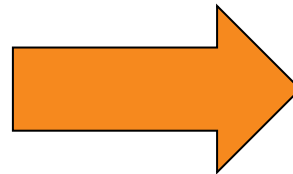
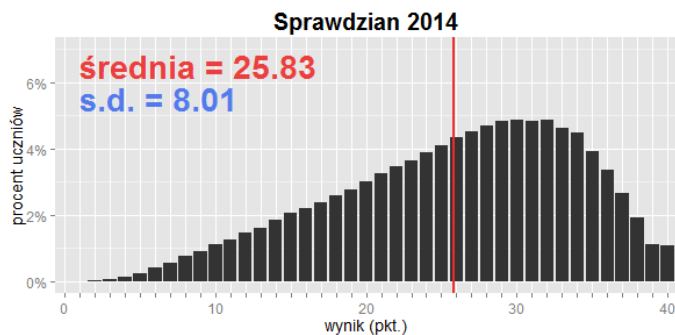
# Zakładany kształt rozkładu wyników testu...

... a faktyczny rozkład wyników.



W praktyce bywa różnie 😊

Możemy jednak rozkład empiryczny spróbować dopasować do założonego





# Popularne skale wyników

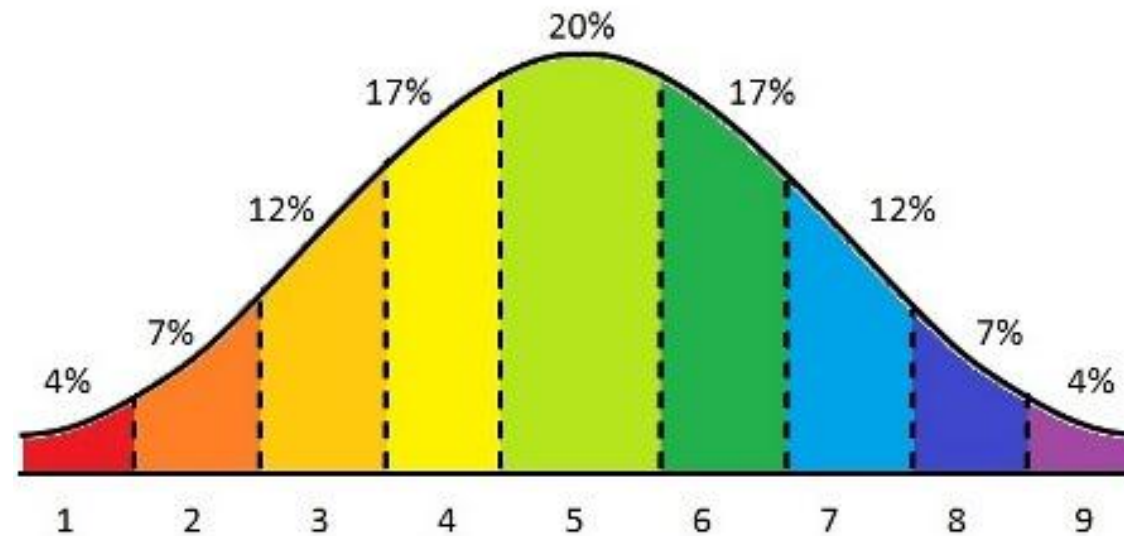
typy skal	Założenia	Przykłady
<b>standardowe</b>	<ul style="list-style-type: none"><li>➤ wyniki mają rozkład normalny</li><li>➤ odchylenie standardowe jako jednostka pomiaru</li></ul>	<ul style="list-style-type: none"><li>➤ skala staninowa</li><li>➤ skala 500/100 (np. PISA)</li><li>➤ skala 100/15</li></ul>
<b>pozycyjne</b>	<ul style="list-style-type: none"><li>➤ odsetek wyników w grupie odniesienia jako jednostka pomiaru</li></ul>	<ul style="list-style-type: none"><li>➤ centylowa</li></ul>



# Skale standardowe

## Skala staninowa

- Dzieli rozkład normalny na 9 przedziałów (staniny)
- Wyniki surowe przypisywane do staninów
- Skala rangowa

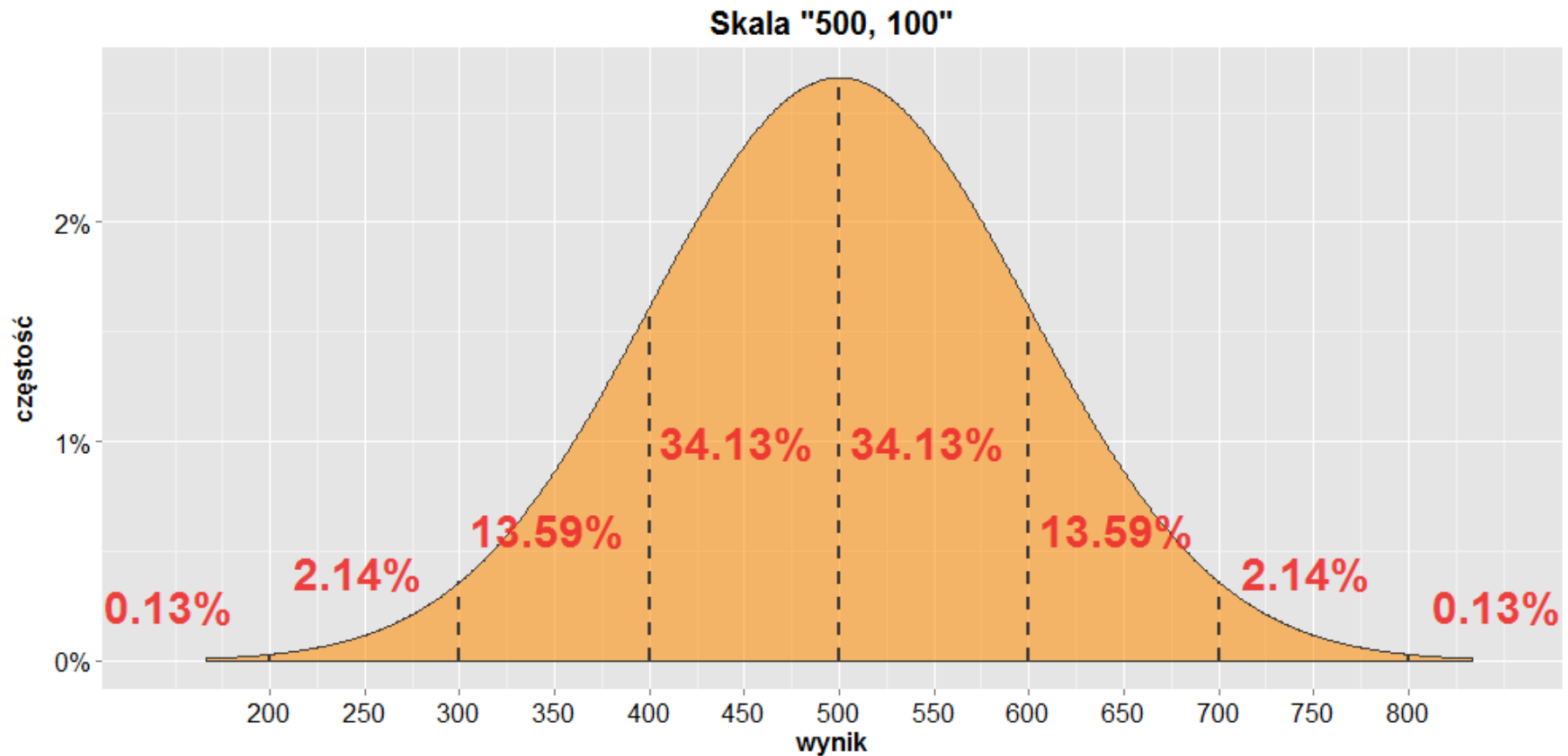




# Skale standardowe

## Skala 500/100

- Wykorzystywana np. w badaniach PISA
- Skala „ciągła”

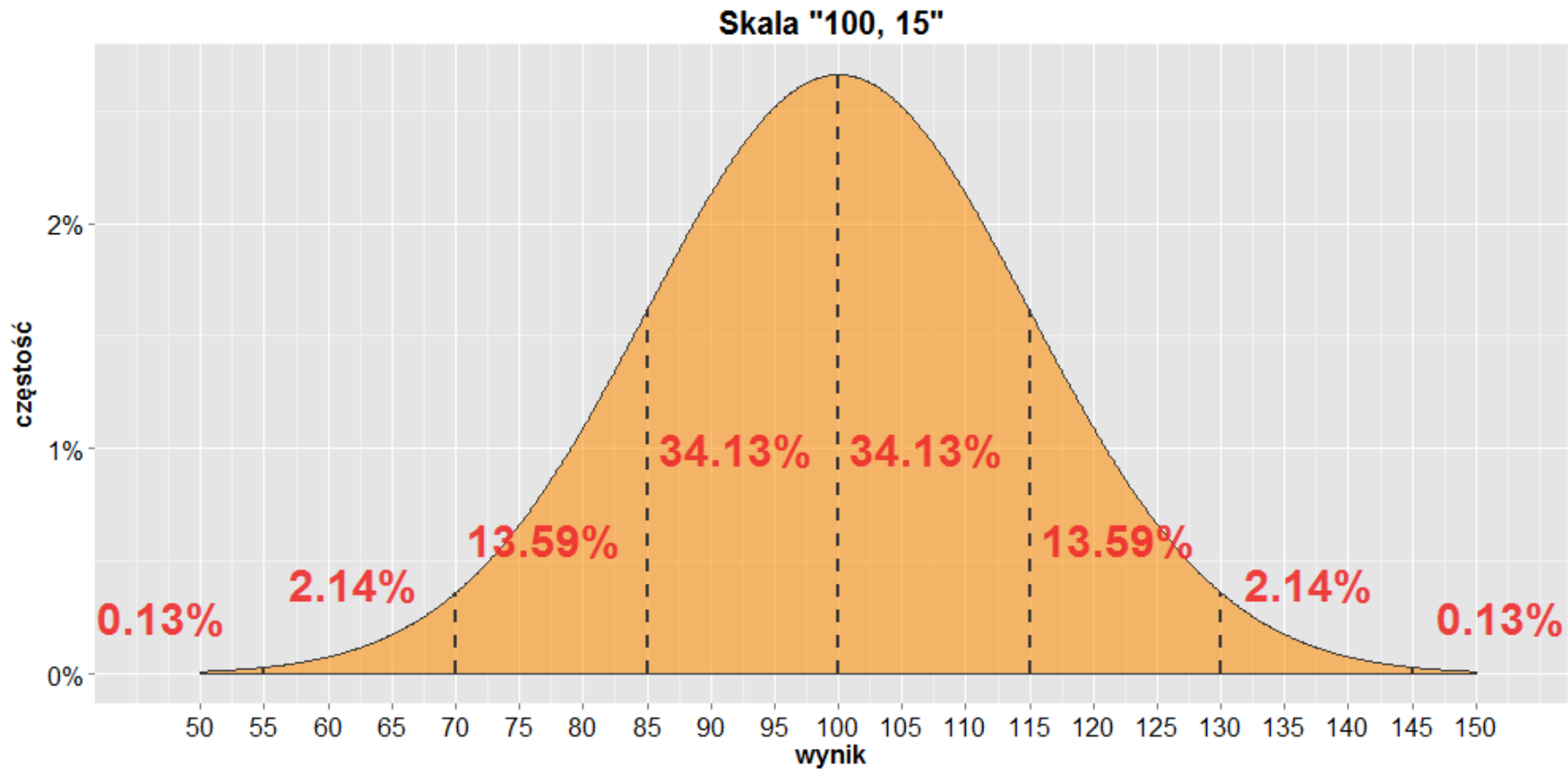




# Skale standardowe

## Skala 100/15

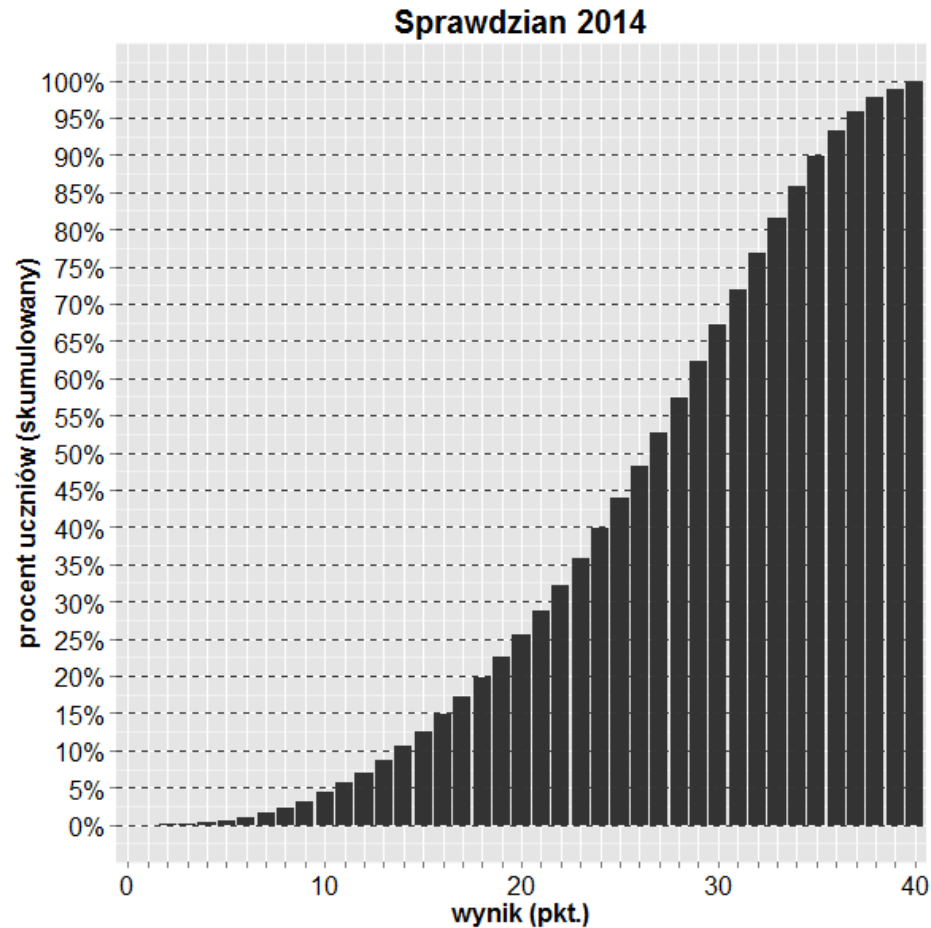
- Tzw. skala „IQ”
- Skala „ciągła”





## Skala centylowa

- Pochodna rozkładu skumulowanego
- Wyniki są przypisywane do rang centylowych
- Przyjmuje się, że jest 99 rang

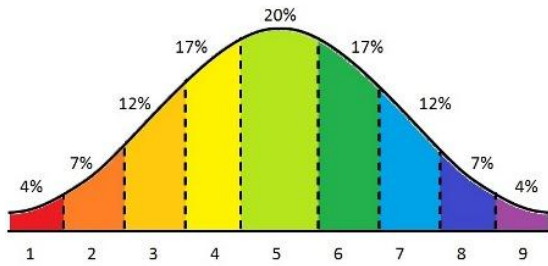
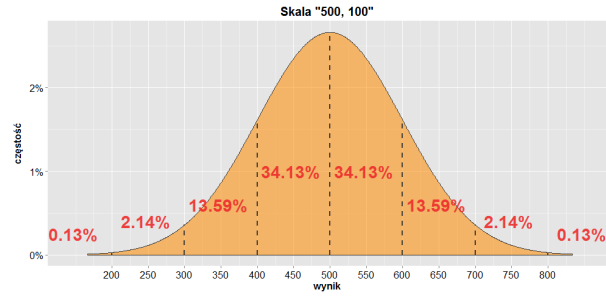




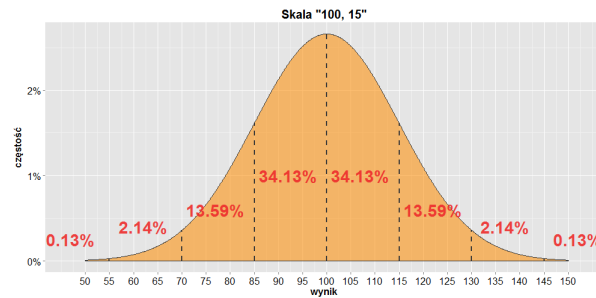


# Którą skalę wybrać?

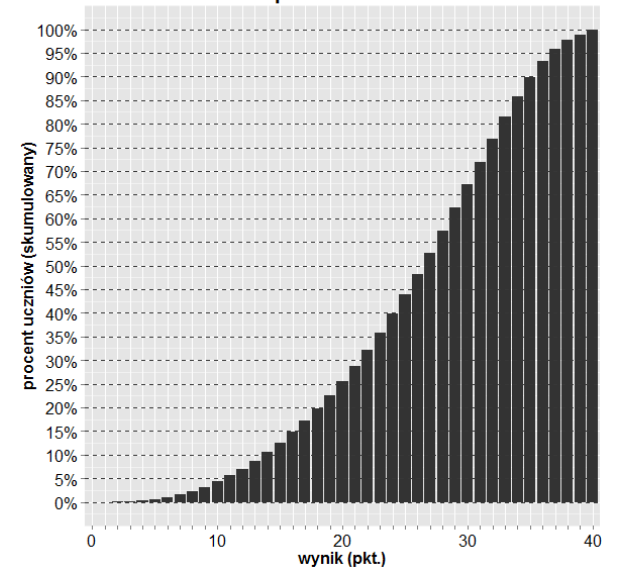
## 500/100



## 100/15



### Sprawdzian 2014





## Pożądane cechy skali wyników



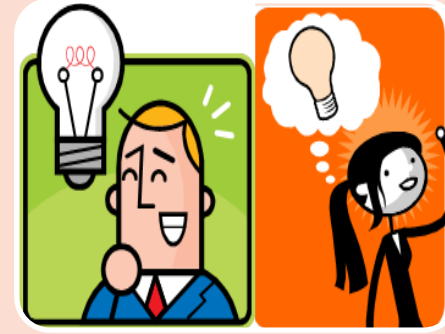
dostateczna  
precyzja

(liczba wyników,  
funkcja testu)



łatwość  
interpretacji  
ilościowej

(wynik  $\Leftrightarrow$  grupa  
odniesienia)



wygoda

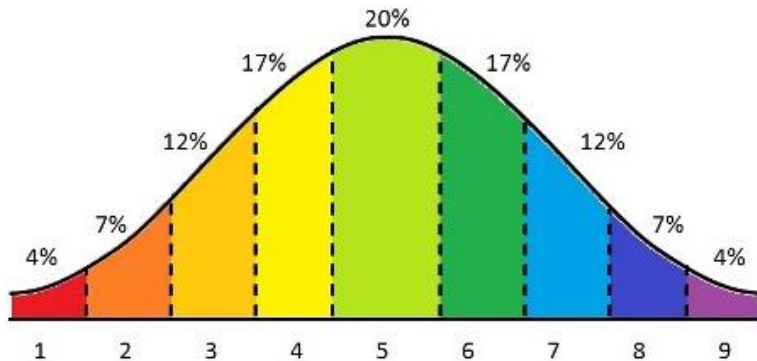
(brak wyników  
ujemnych i  
ułamkowych)



## Którą skalę wybrać?

Problemy ze skalą staninową

- Za „gruboziarnista”
- Utrata informacji: 41 wyników → 9 wartości
- W obrębie stanina nie rozróżniamy wyników



stanin	wynik Sprawdzian 2014
1	0 - 9
2	11 - 14
3	15 - 19
4	20 - 24
5	25 - 28
6	29 - 32
7	33 - 34
8	35 - 37
9	38 - 40

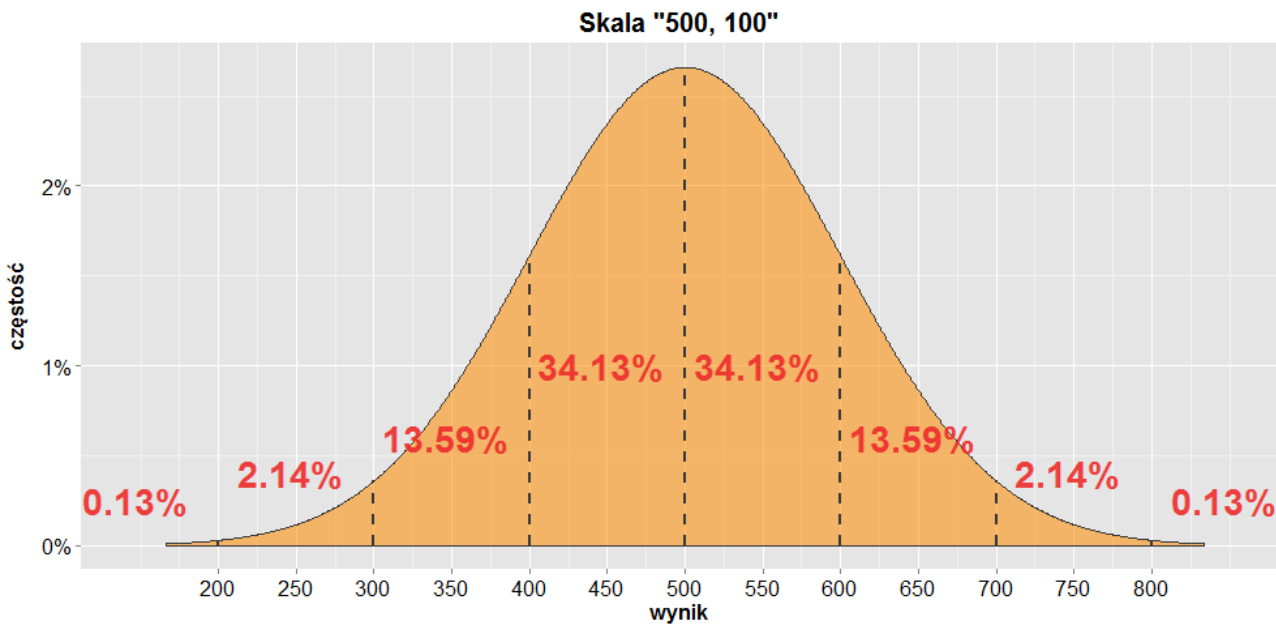
Tu jest  
dwadzieścia  
kategorii  
punktowych



# Którą skalę wybrać?

Problemy ze skalą 500/100

- Za „drobnoziarnista”
- Złudzenie precyzji (zwłaszcza przy krótkich testach)
- Skala wyników dziurawa



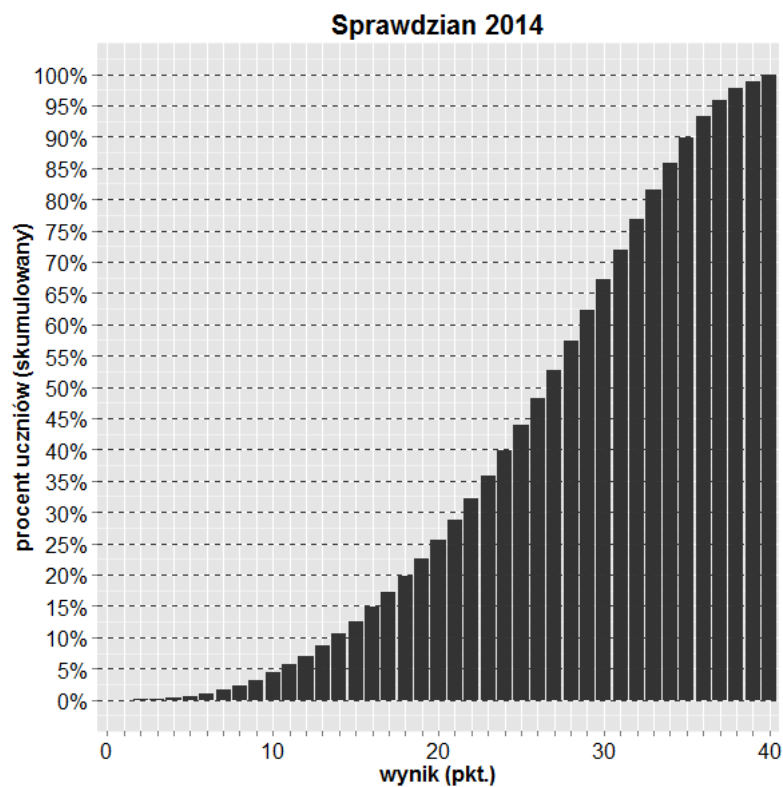
wynik Sprawdzian 2014	wynik na skali 500/100
...	...
5	245
6	266
7	284
...	...
20	434
21	444
22	454
...	...



# Którą skalę wybrać?

Problemy ze skalą centylową

- Ryzyko złączenia się kategorii punktowych
- Ryzyko dziur w rangach (przy krótkim teście)
- Statystyka lubi rozkład normalny 😊



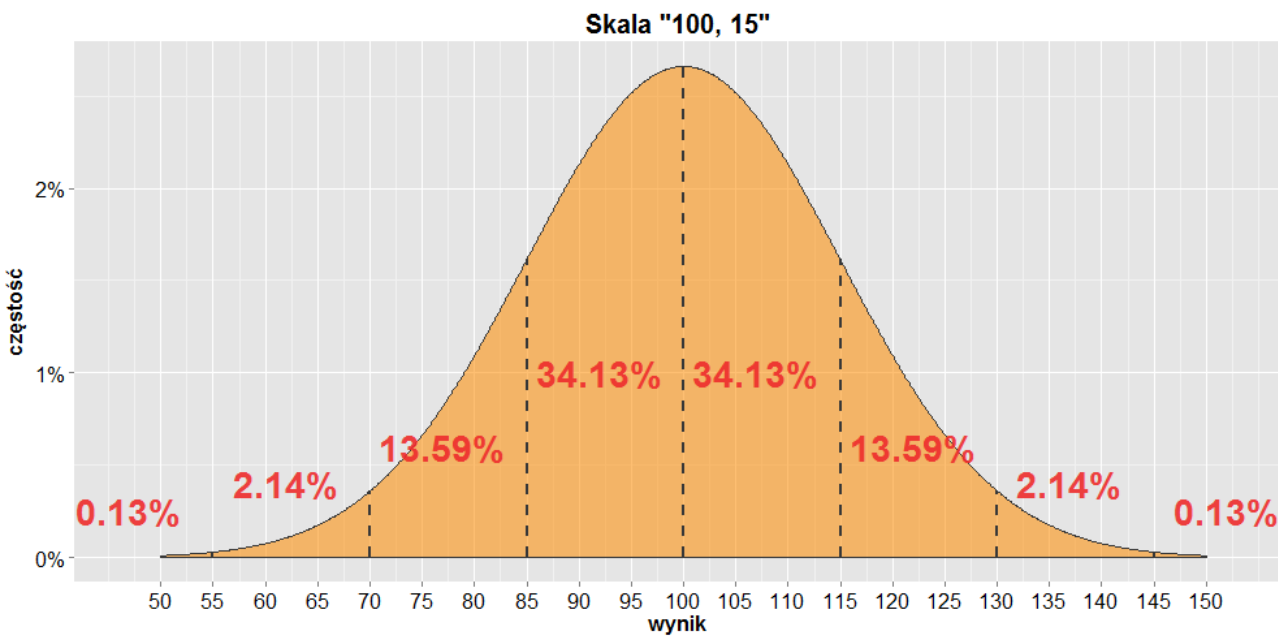
wynik Sprawdzian 2014	Percentyl (max. ranga)
...	...
5	1
6	1
7	1
...	...
20	25
21	28
22	32
...	...



# Którą skalę wybrać?

Skala 100/15

- Dobra „gęstość” wyników
- Sensowny „zapas” precyzji (dla dłuższych testów)
- Nasz wybór 😊!



wynik Sprawdzian 2014	wynik na skali 100/15
...	...
5	62
6	65
7	68
...	...
20	90
21	92
22	93
...	...



# Jak przełożyć wyniki surowe na skalę 100/15

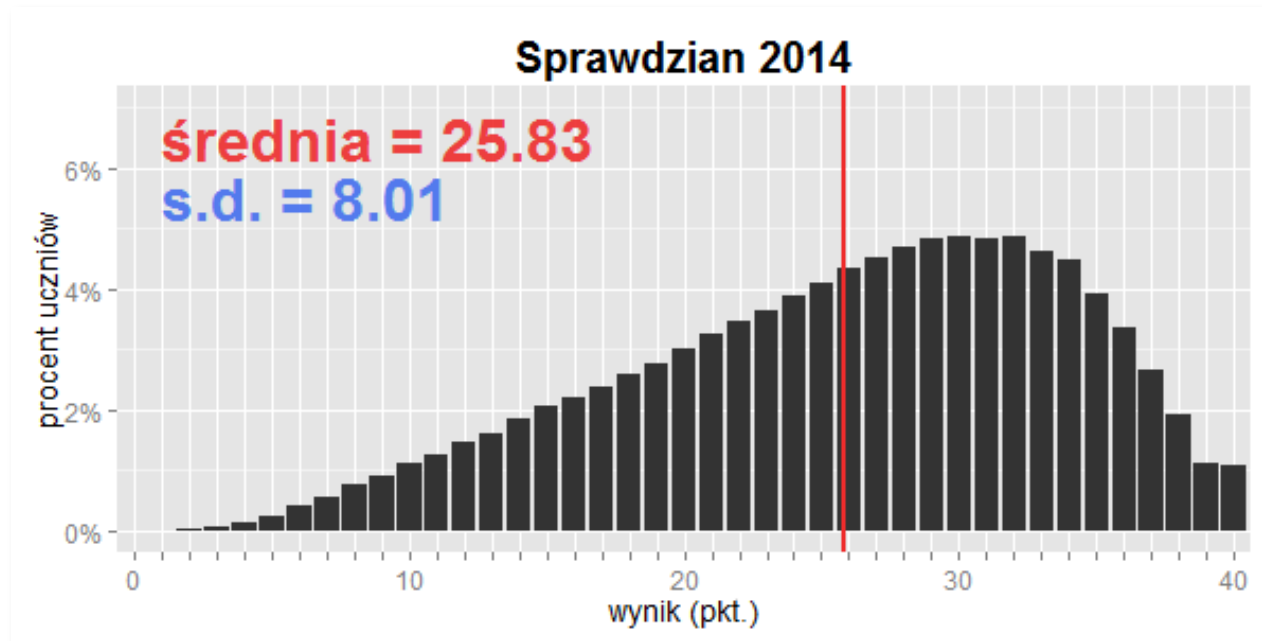
metodą rozkładu skumulowanego  
w czterech prostych krokach





# Jak przełożyć wyniki surowe na skalę 100/15

metodą rozkładu skumulowanego  
w czterech prostych krokach



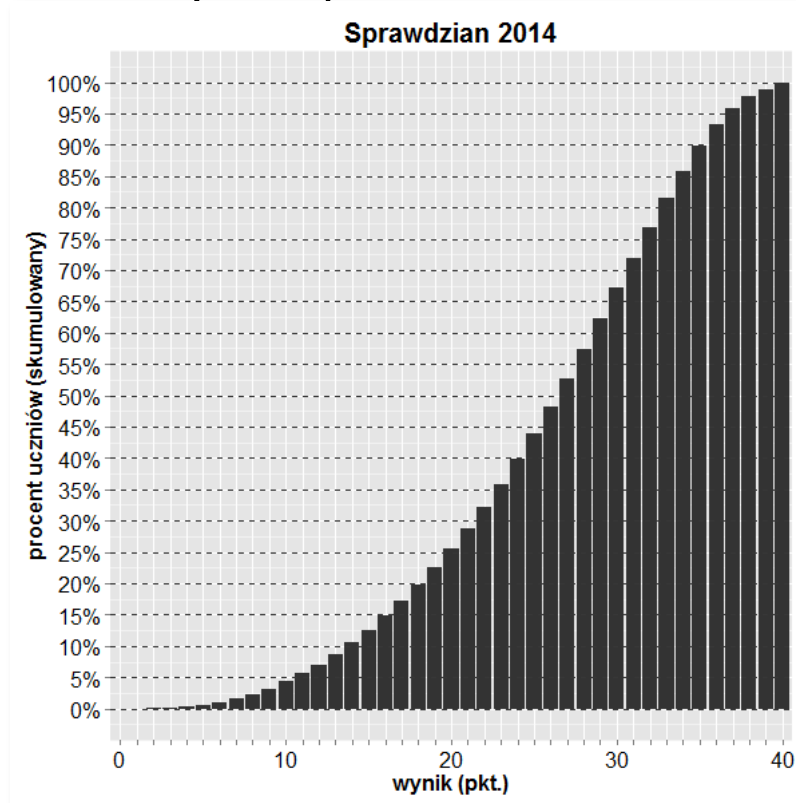
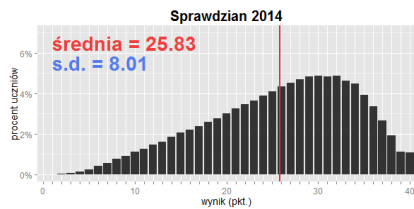
Bierzemy rozkład  
(procentowy)  
wyników surowych





# Jak przełożyć wyniki surowe na skalę 100/15

metodą rozkładu skumulowanego  
w czterech prostych krokach



Bierzemy rozkład  
wyników surowych  
(procentowy)

Przekształcamy go  
na rozkład  
skumulowany



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE  entuzjaści  
edukacji

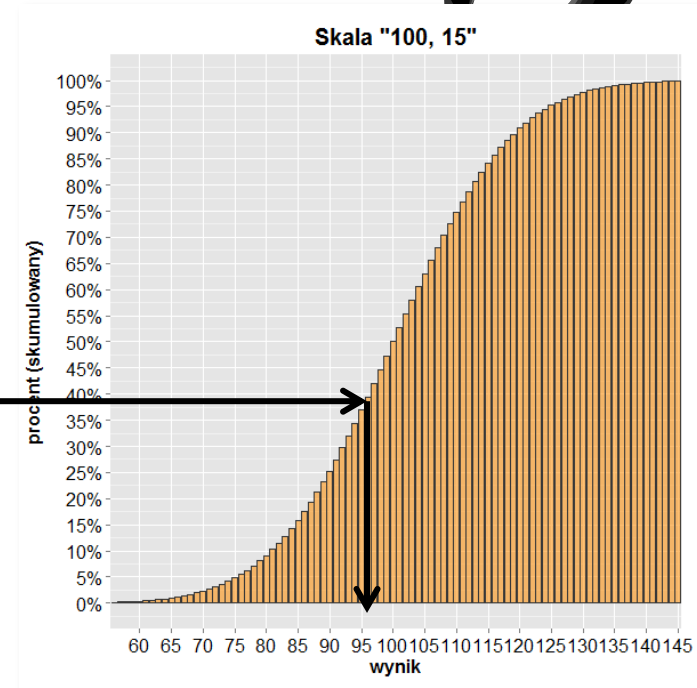
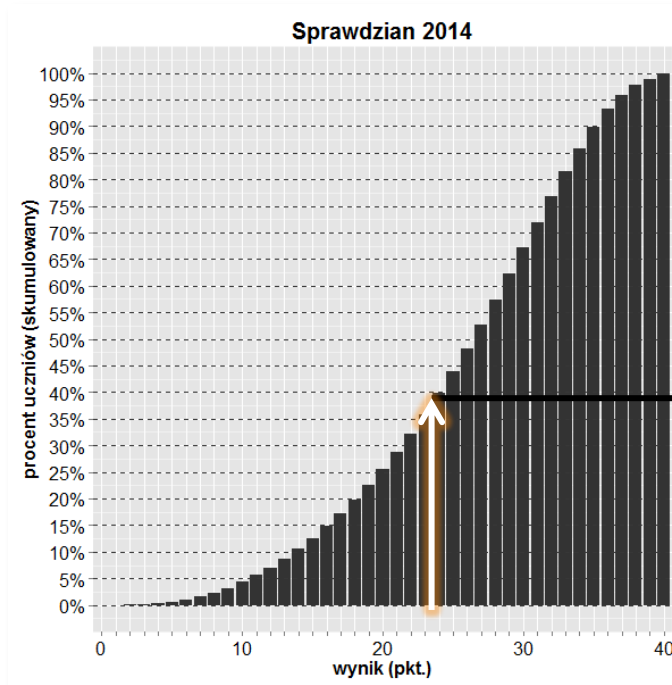
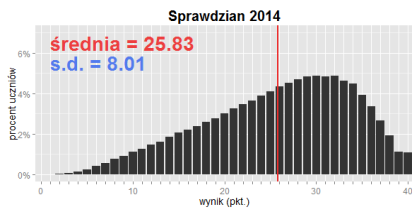
UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOLECZNY





# Jak przełożyć wyniki surowe na skalę 100/15

metodą rozkładu skumulowanego  
w czterech prostych krokach



Bierzemy rozkład  
wyników surowych  
(procentowy)

Przekształcamy go  
na rozkład  
skumulowany

Porównujemy z  
rozkładem  
skumulowanym dla  
skali 100/15



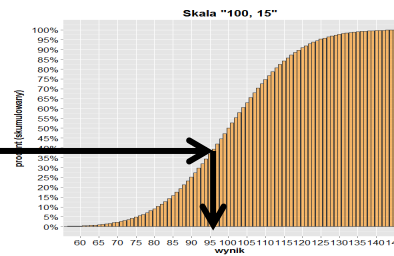
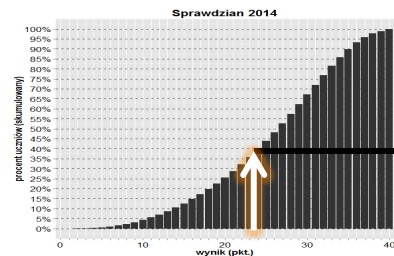
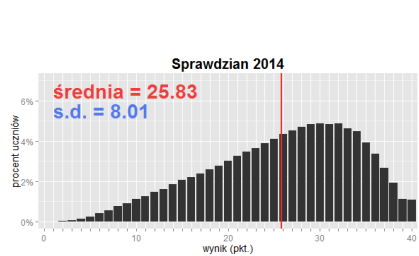
KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

IBE  entuzjaści  
edukacji

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY 



# Jak przełożyć wyniki surowe na skalę 100/15 metodą rozkładu skumulowanego



wynik Sprawdzian 2014	wynik na skali 100/15
...	...
5	62
6	65
7	68
...	...

Bierzemy rozkład  
wyników surowych  
(procentowy)

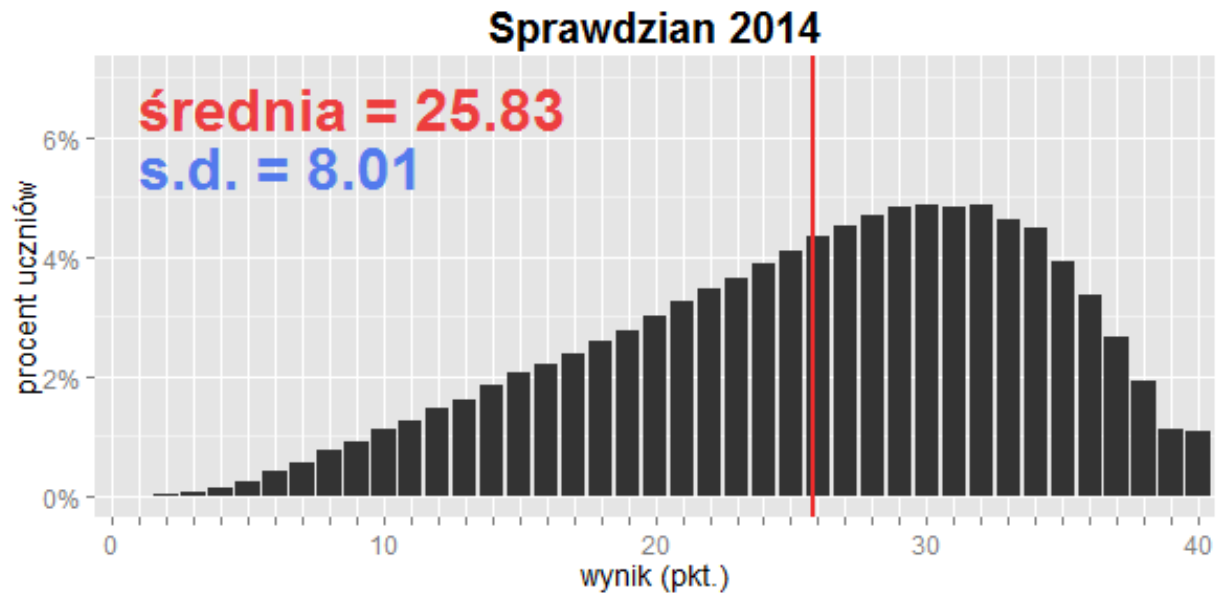
Przekształcamy go  
na rozkład  
skumulowany

Porównujemy z  
rozkładem  
skumulowanym dla  
skali 100/15

I otrzymujemy  
wyniki na skali  
100/15



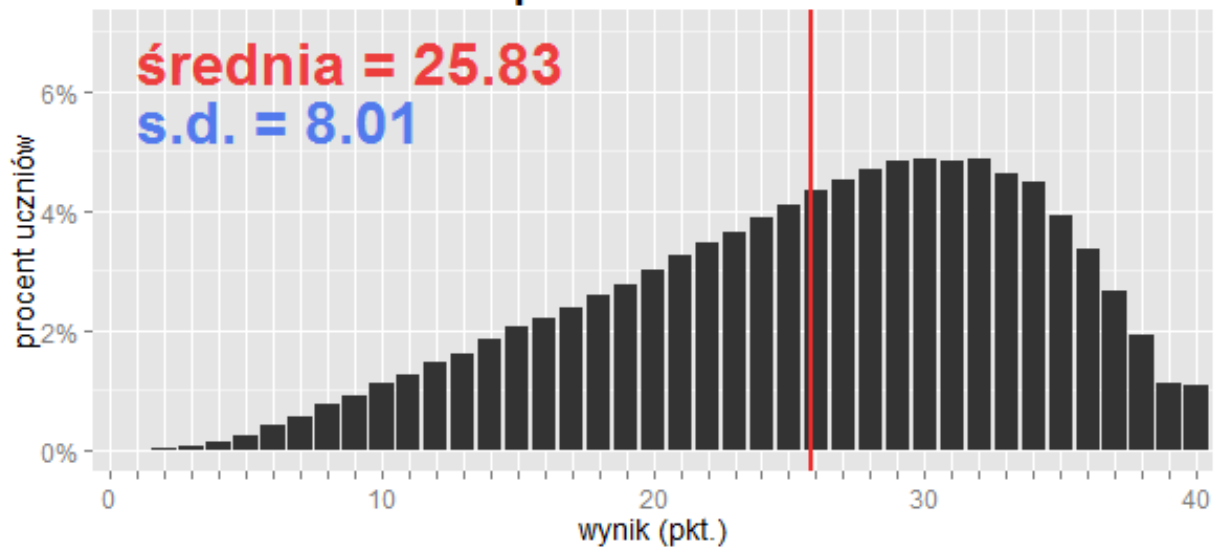
## Przed...



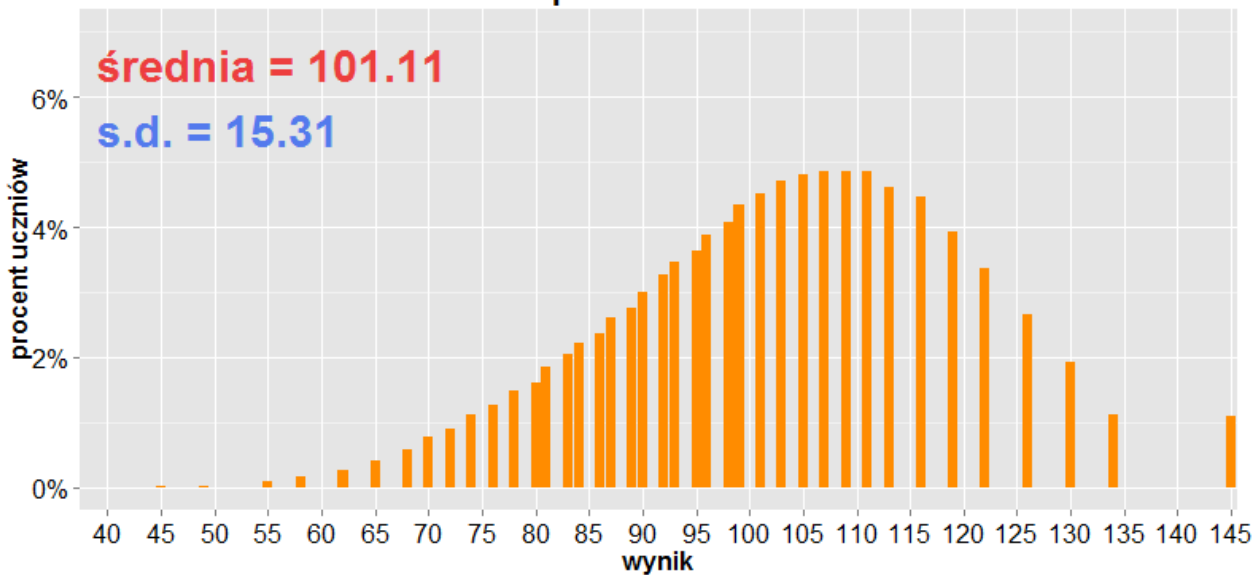


# Przed i po

### Sprawdzian 2014

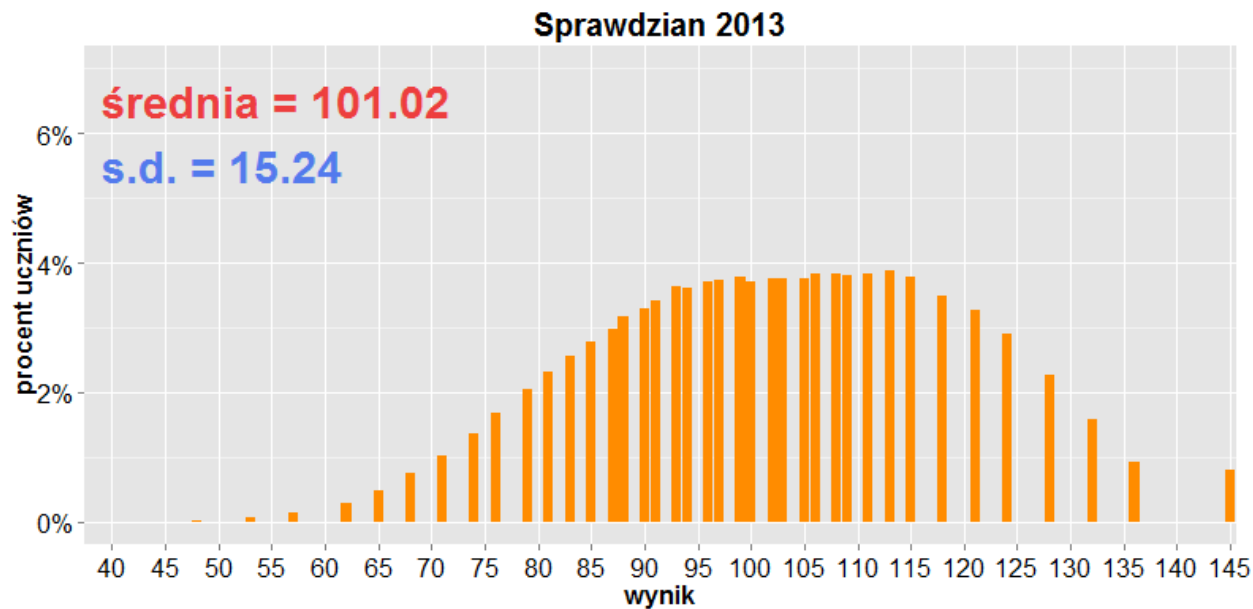
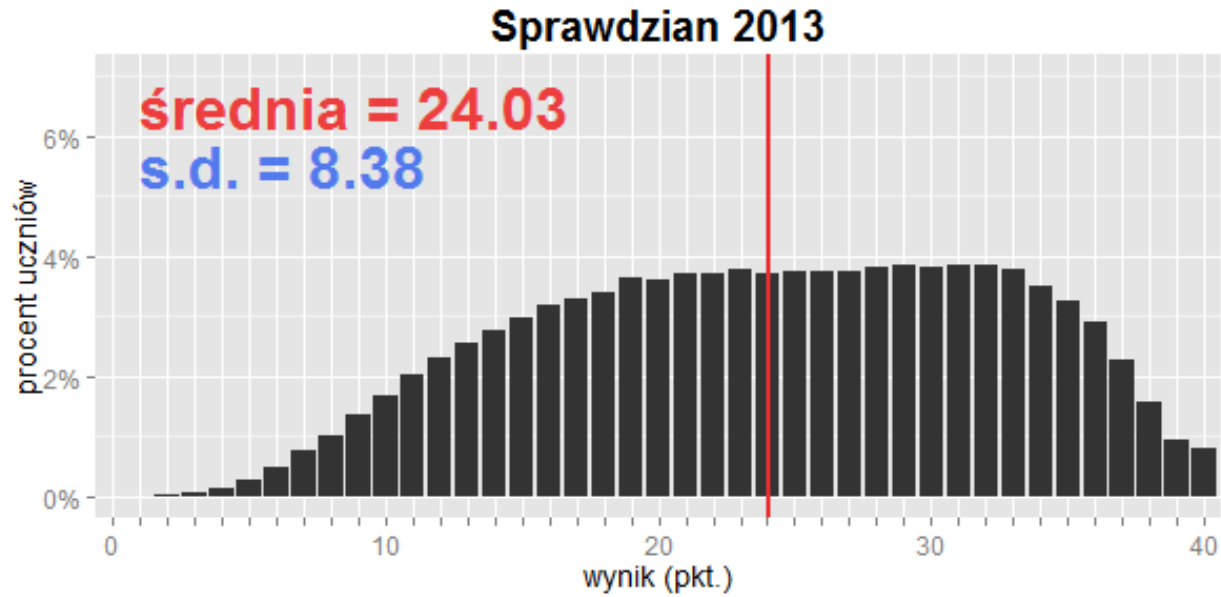


### Sprawdzian 2014





# Sprawdzian 2013





## Podsumowanie

- Wyniki mają znany rozkład w grupie odniesienia.
- Wynik na skali łatwo porównać do rozkładu w grupie odniesienia (ile jest wyników poniżej/powyżej danego wyniku).
- Różnice w wynikach indywidualnych na skali 100/15 są sensownie interpretowalne. Nie ma potrzeby wykorzystania wyników ułamkowych.
- Może „obsłużyć” krótsze i dłuższe testy.



## Podsumowanie

- Wyniki mają znany rozkład w grupie odniesienia.
  - Wynik na skali łatwo porównać do rozkładu w grupie odniesienia (ile jest wyników poniżej/powyżej danego wyniku).
  - Różnice w wynikach indywidualnych na skali 100/15 są sensownie interpretowalne. Nie ma potrzeby wykorzystania wyników ułamkowych.
  - Może „obsłużyć” krótsze i dłuższe testy.
- 
- Nie naprawi jednak wad testów (efekt sufitowy, podłogowy, „dziwne” rozkłady wyników).





## Gdzie jest wykorzystywana skala 100/15 ?

- Porównywalne Wyniki Egzaminów
  - <http://pwe.ibe.edu.pl>
- Kalkulator EWD SP, wskaźniki EWD dla gimnazjów i szkół maturalnych
  - <http://ewd.edu.pl>
- Testy wykorzystane w badaniu SUEK (TOS3 i TOS6)
- OBUT
- Test umiejętności na starcie szkolnym (TUNSS)
- Niektóre testy psychologiczne



*entuzjaści edukacji*

**Dziękuję za uwagę!**

**„Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego”**

*Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego*

**Instytut Badań Edukacyjnych**  
ul. Górczewska 8, 01-180 Warszawa  
tel.: (22) 241 71 00, e-mail: [ibe@ibe.edu.pl](mailto:ibe@ibe.edu.pl)