Autor: **Krzysztof Dąbrowski 293101** 29 maja 2020

POSI – Laboratoria C2

Badanie metod skalowania

# Opis zadania

Celem laboratorium jest zbadanie wpływy różnych metod skalowania liniowego na wartości funkcji ewaluacji w algorytmie genetycznym. W tym celu badane zostaną po dwa zestawy danych dla dówch różnych funkcji.

Interaktywny zeszyt Jupyter Notebook z obliczeniami i analizą danych wykonanymi na potrzeby zadania jest dostępny pod adresem <https://github.com/SiwyKrzysiek/POSI-C2>

# Badanie dla funkcji De Jonga

Funkcja De Jonga nazywana również paraboloidą jest zdefiniowana poniższym równaniem.

Przy pomocy dwóch zbiorów danych testowych została zbadana metoda skalowania liniowego oraz skalowania σ-odcinającego.

## Mały zbiór danych

Pierwszy badany zbiór składał się z 20 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.

A close up of a piece of paper

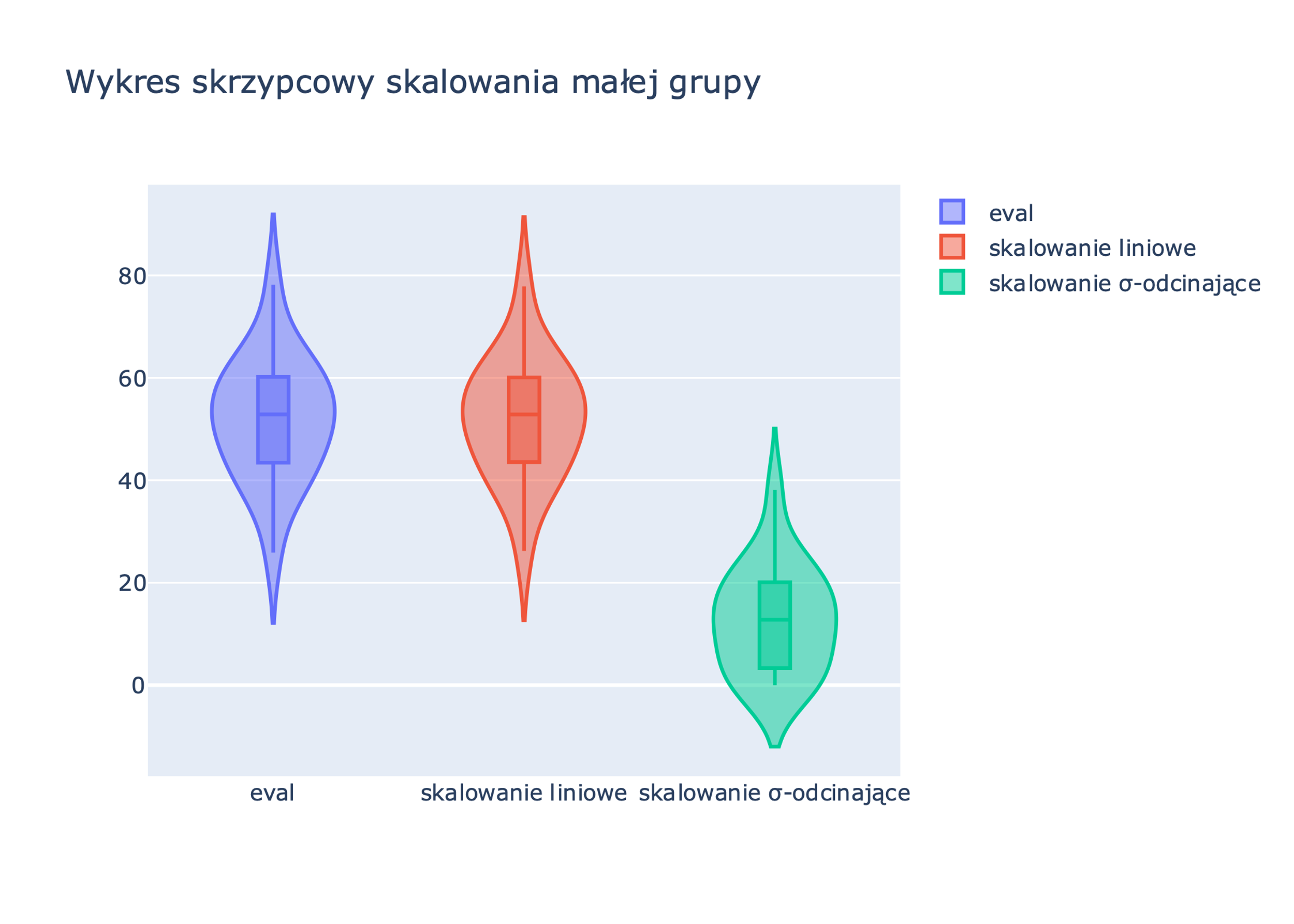
Description automatically generated

Rysunek 1 - wartości małej grupy dla paraboloidy

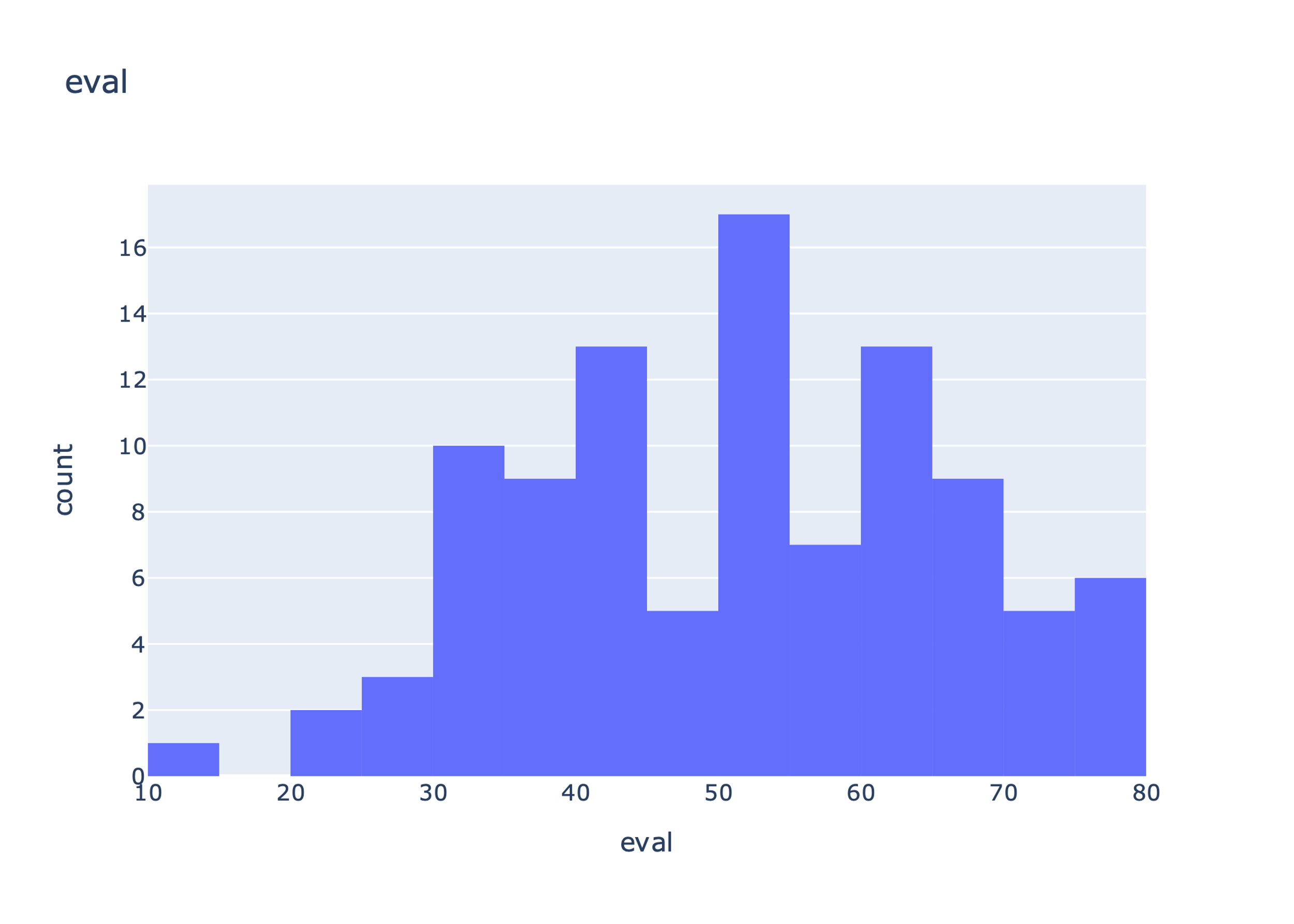
A screenshot of a cell phone

Description automatically generated

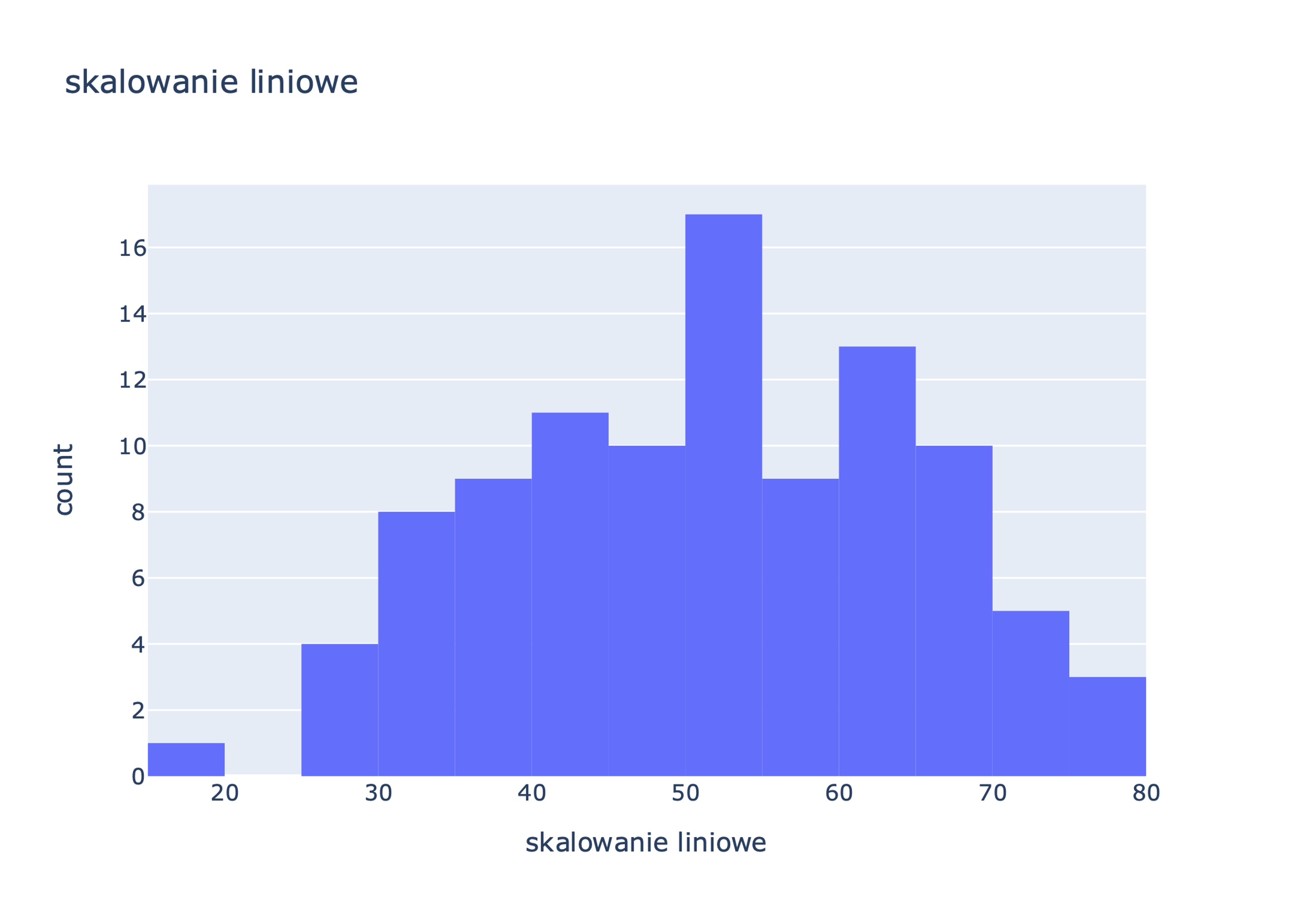
Rysunek 2 - Parametry statystyczne skalowania małej grupy dla paraboloidy



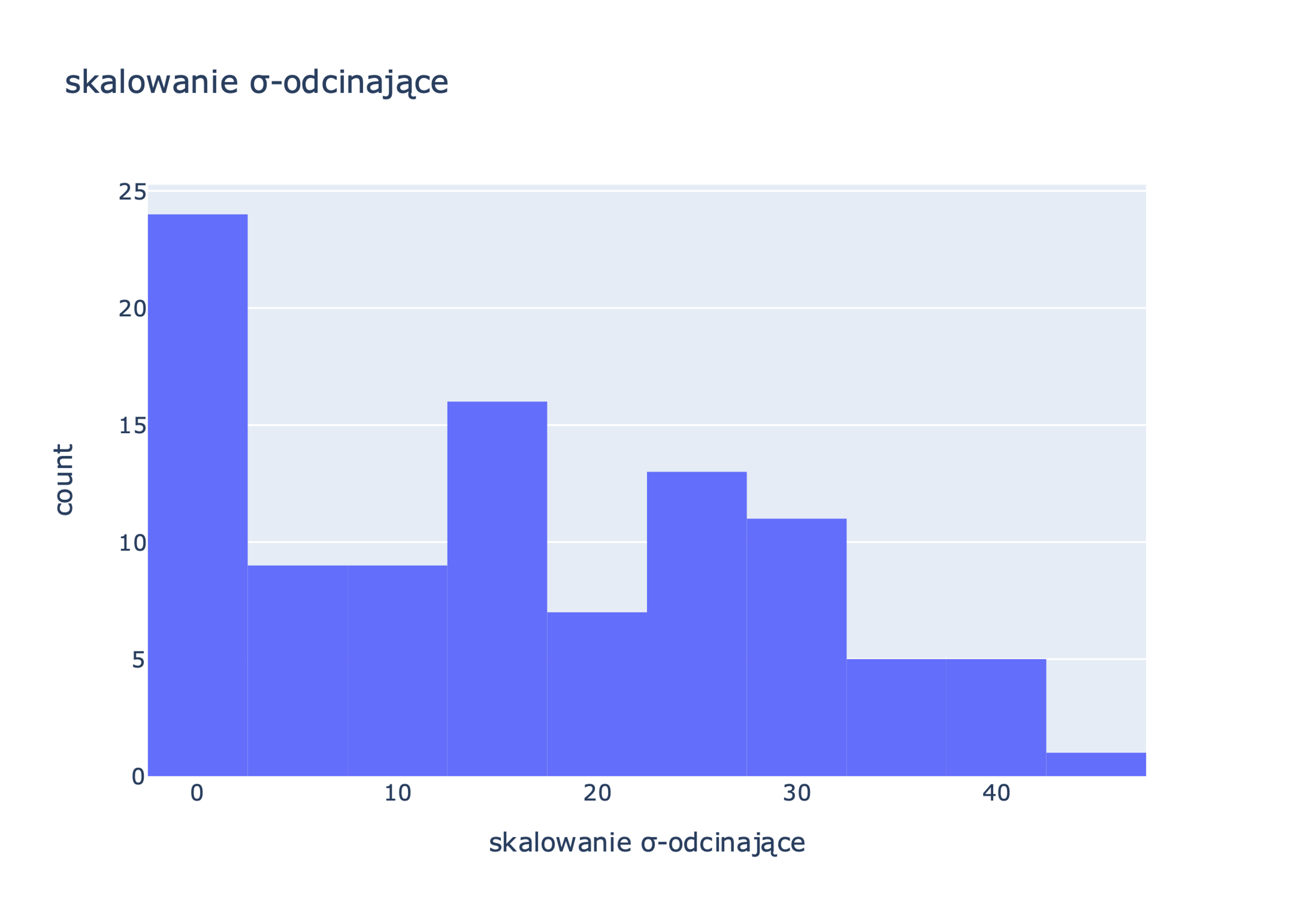
Rysunek 3 - Wykres skrzypcowy skalowania małej grupy dla paraboloidy



Rysunek 4 - histogram ewaluacji małej grupy dla paraboloidy



Rysunek 5 - histogram skalowania liniowego małej grupy dla paraboloidy



Rysunek 6 - histogram skalowania σ-odcinającego małej grupy dla paraboloidy

## Dużego zbiór danych

Pierwszy badany zbiór składał się z 100 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.

A screenshot of a cell phone

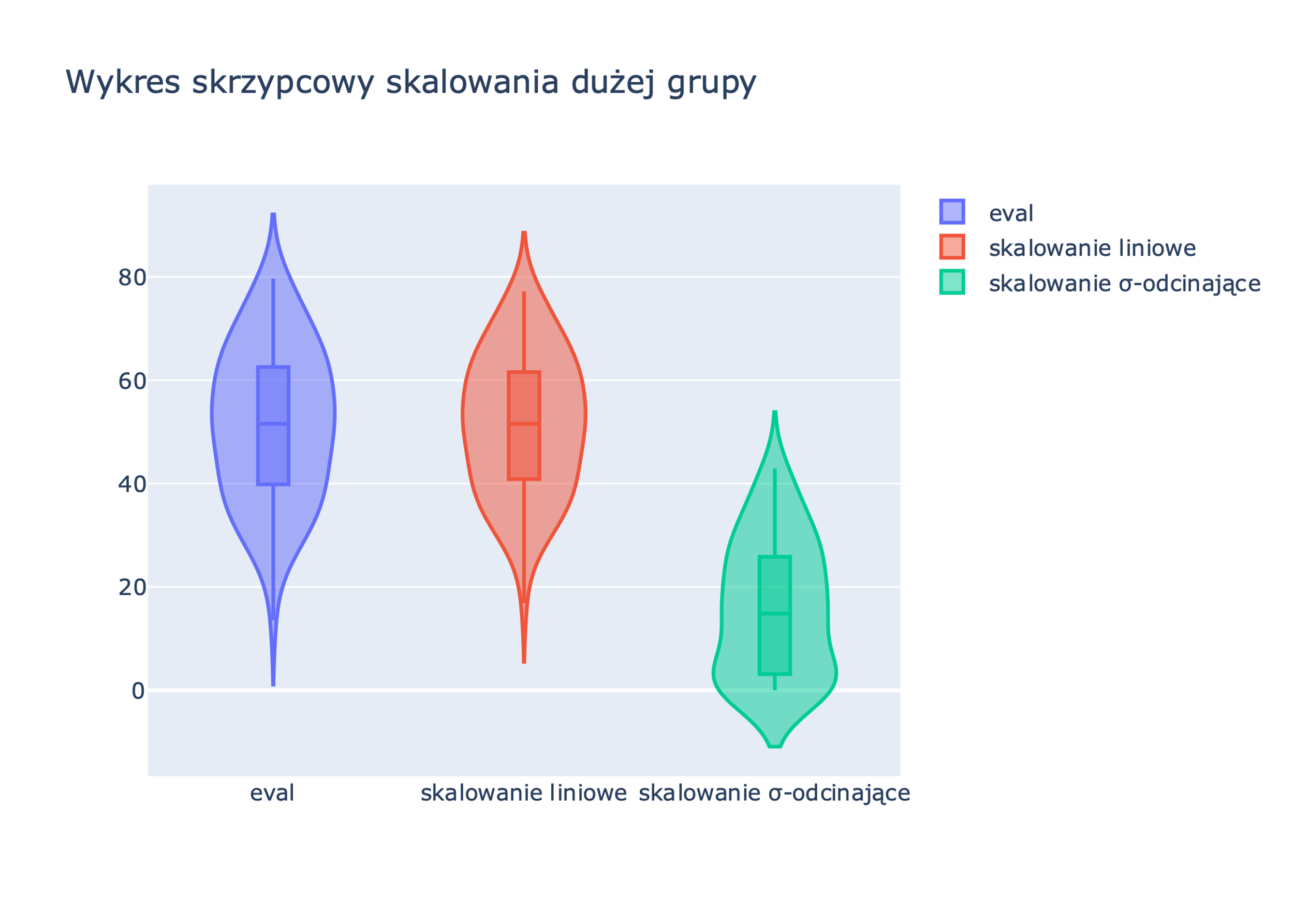
Description automatically generated

Rysunek 7 - Przykładowe wartości dużej grupy dla paraboloidy

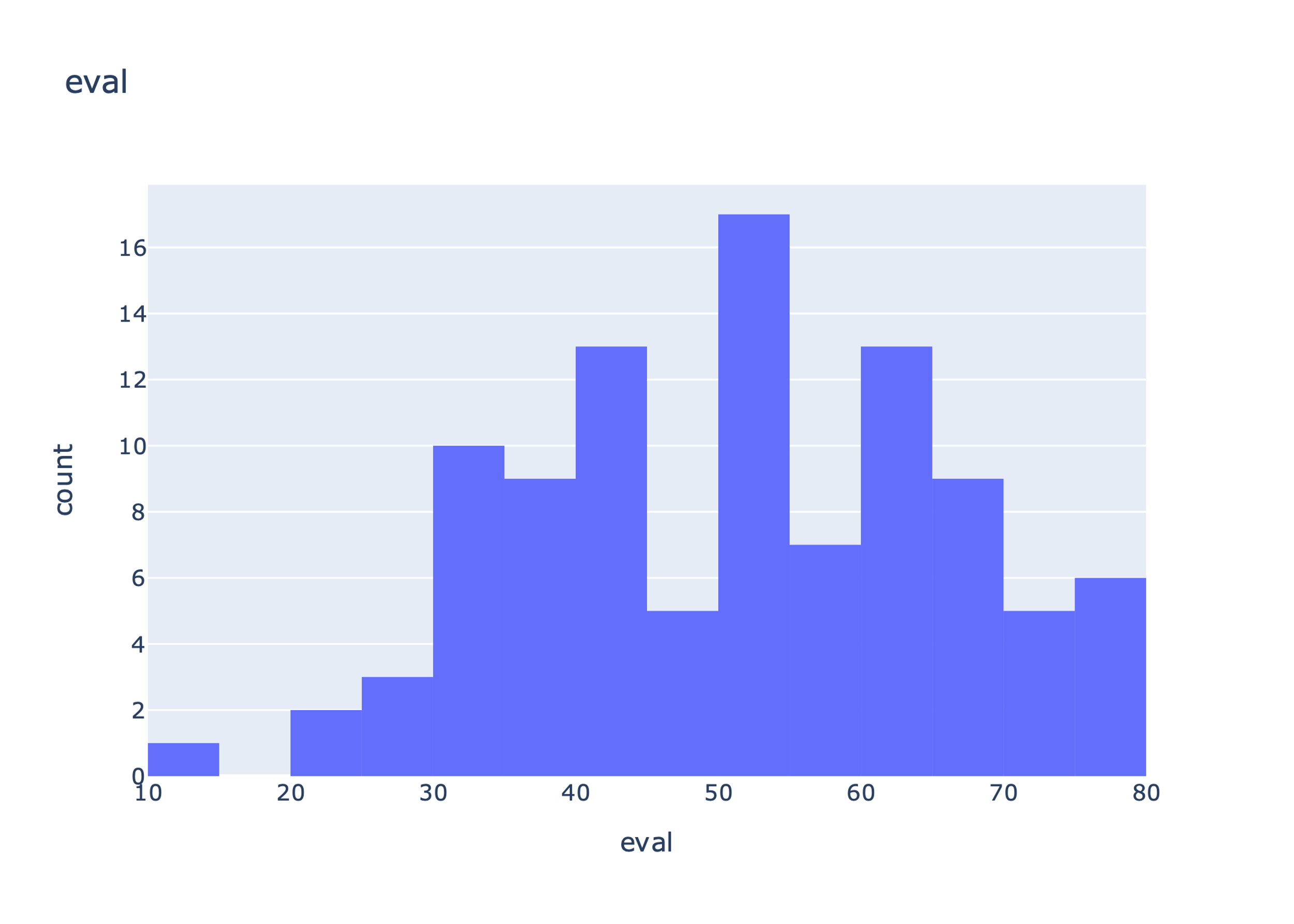
A screenshot of a cell phone

Description automatically generated

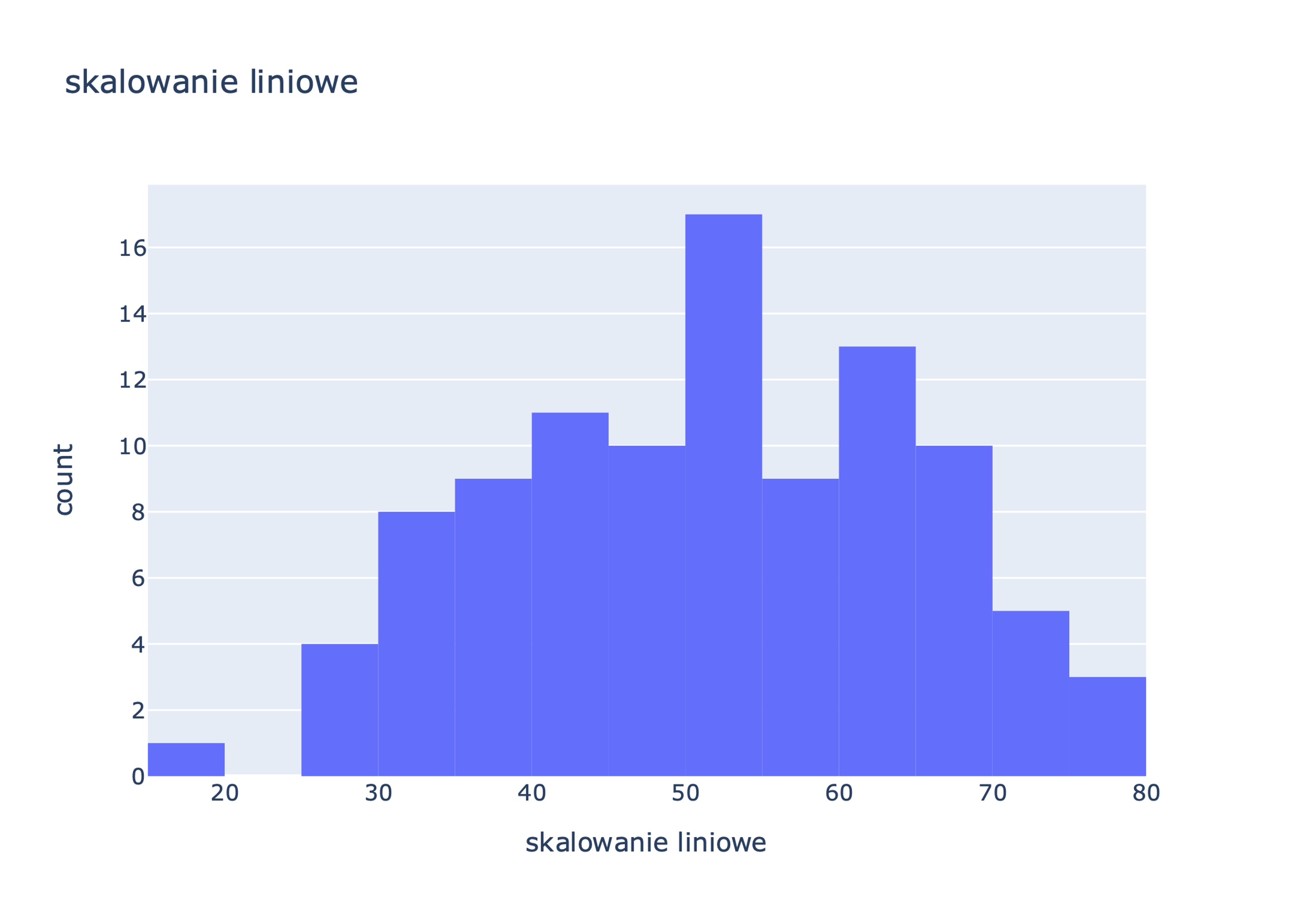
Rysunek 8 - Parametry statystyczne skalowania dużej grupy dla paraboloidy



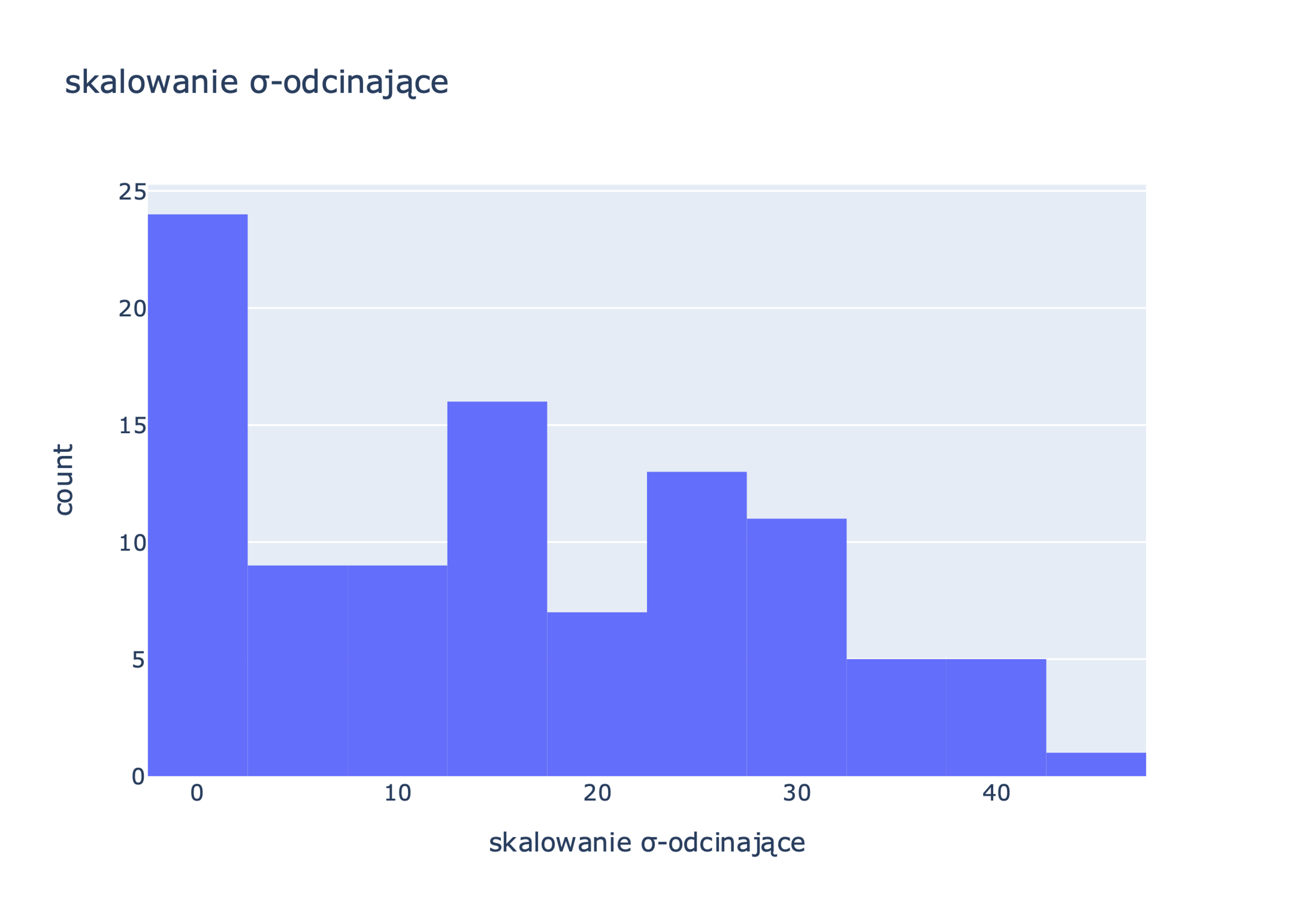
Rysunek 9 - Wykres skrzypcowy skalowania małej dużej grupy dla paraboloidy



Rysunek 10 - histogram ewaluacji dużej grupy dla paraboloidy



Rysunek 11 - histogram skalowania liniowego dużej grupy dla paraboloidy



Rysunek 12 - histogram skalowania σ-odcinającego dużej grupy dla paraboloidy

## Wnioski

Skalowanie liniowe w niewielkim stopniu modyfikuje parametry statystyczne danych.

Na podstawie histogramów można zaobserwować, że skalowanie liniowe niejako wypełnia luki wygładzając histogram wartości.

Skalowanie σ-odcinające zmniejsza wartości każdego elementu. Jednak nie wpływa zbytnio na odchylenie standardowe danych.

Na wykresach skrzypcowych oraz na histogramach można zaobserwować, że skalowanie σ-odcinające grupuje sporą część danych w okolicy początkowych wartości. Może to wynikać z mapowania wszystkich ujemnych wartości na 0.

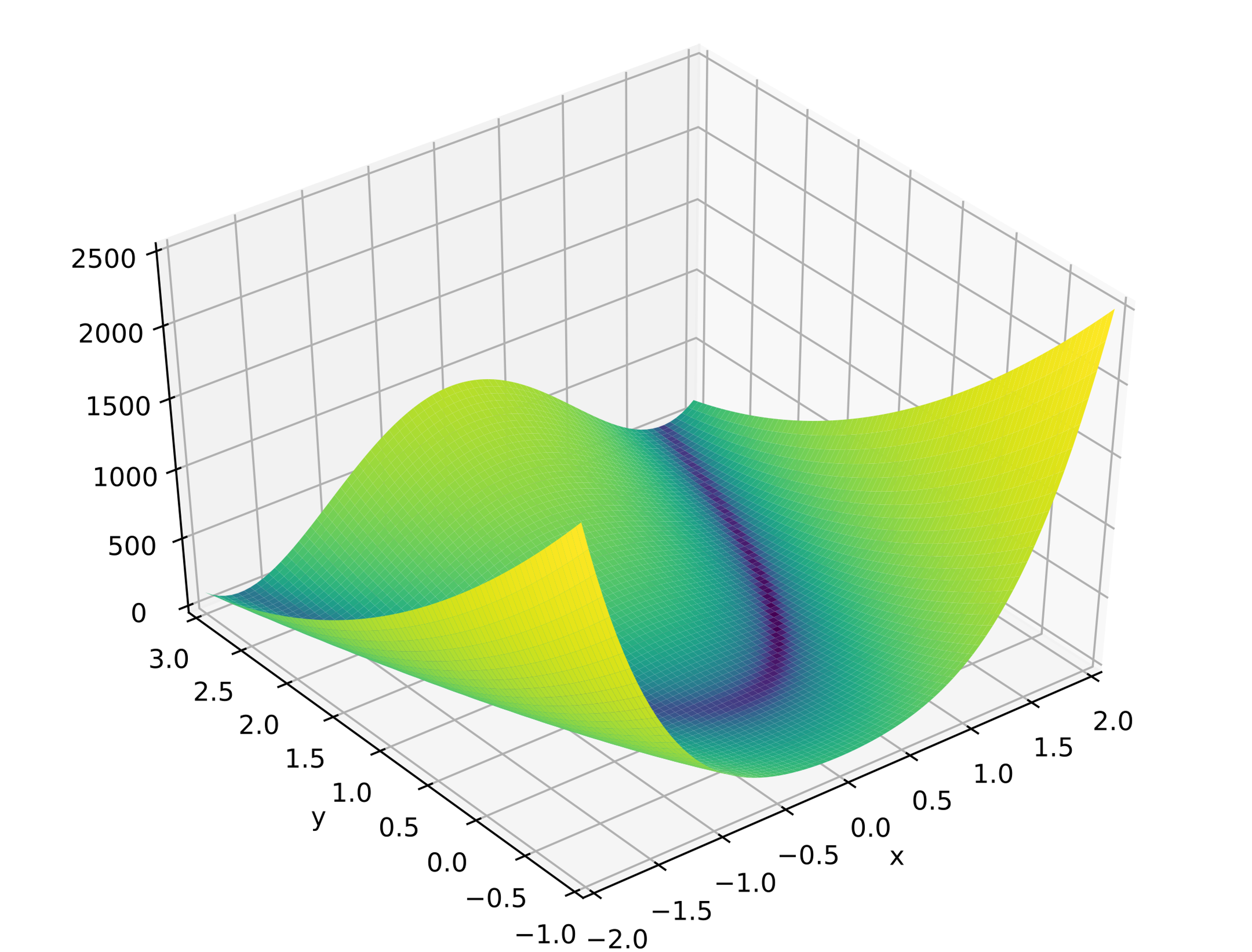
Zmniejszenie wartości wszystkich wartości w skalowaniu σ-odcinającym wynika bezpośrednio ze wzoru, który zakłada odejmowanie stałej od każdej wartości danych.

# Badanie funkcji siodła Rosenbrocka

Funkcja Rosenbrocka jest parametryzowaną funkcją dwóch zmiennych. Jej wykres przypomina kanion wyżłobiony przez górski strumień. Funkcja ta jest często wykorzystywana przy testowaniu algorytmów optymalizacyjnych. Bardzo łatwo znaleźć jest jej dolinę, jednak dojście do minimum globalnego jest nietrywialne.

Funkcja dana jest poniższym wzorem:

Typowo funkcja ta jest stosowana z parametrami a = 1 i b = 100. Takie parametry zostały zastosowane w badaniach laboratoryjnych.



Rysunek 13 - Wykres funkcji Rosenbrocka dla standardowych parametrów  
Źródło: Wikipedia

Przy pomocy dwóch zbiorów danych testowych została zbadana metoda skalowania rankingowa linowa oraz skalowania logarytmicznego.

## Mały zbiór danych

Pierwszy badany zbiór składał się z 20 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.

A screenshot of a cell phone

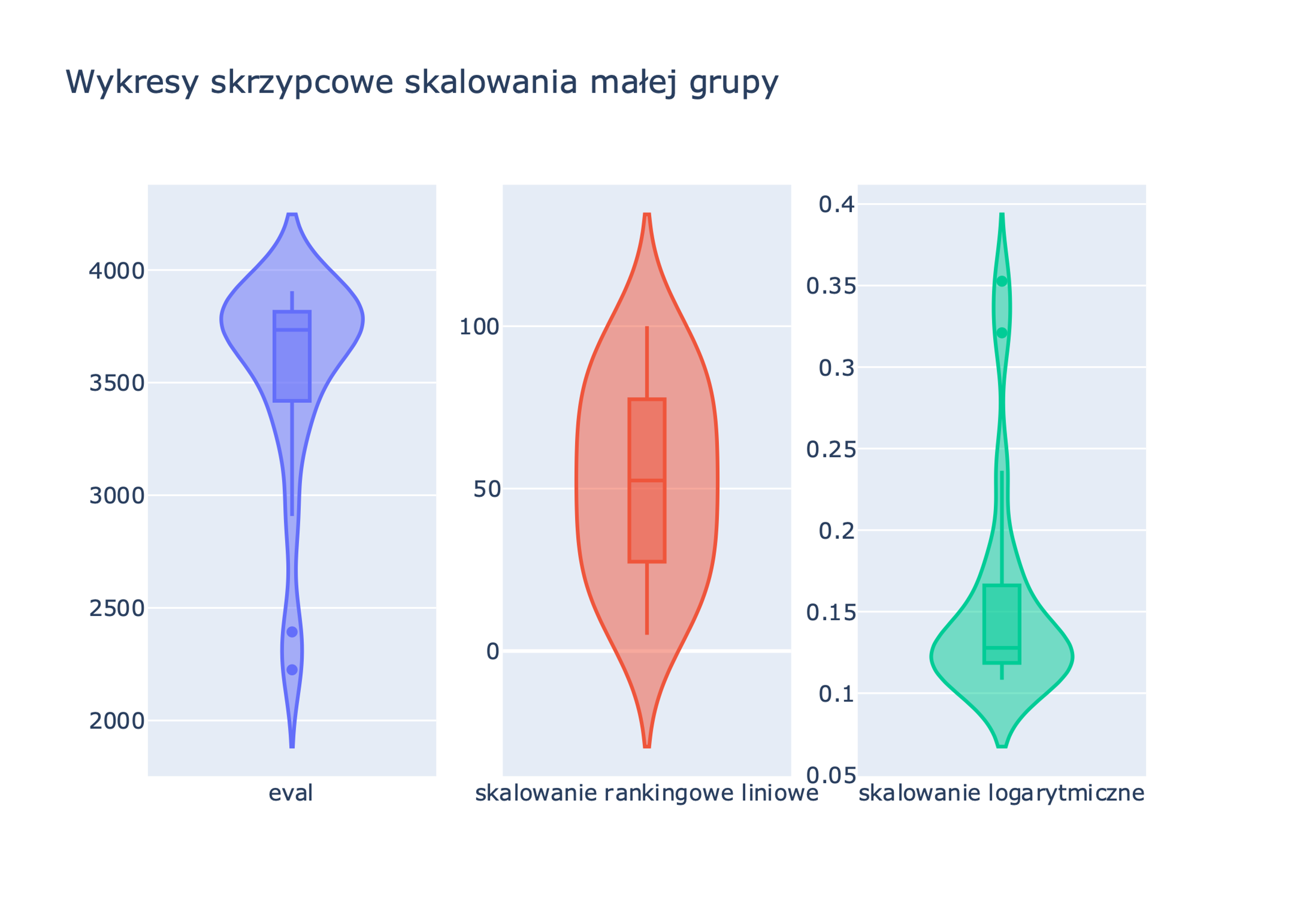
Description automatically generated

Rysunek 14 - wartości małej grupy dla siodła Rosenbrocka

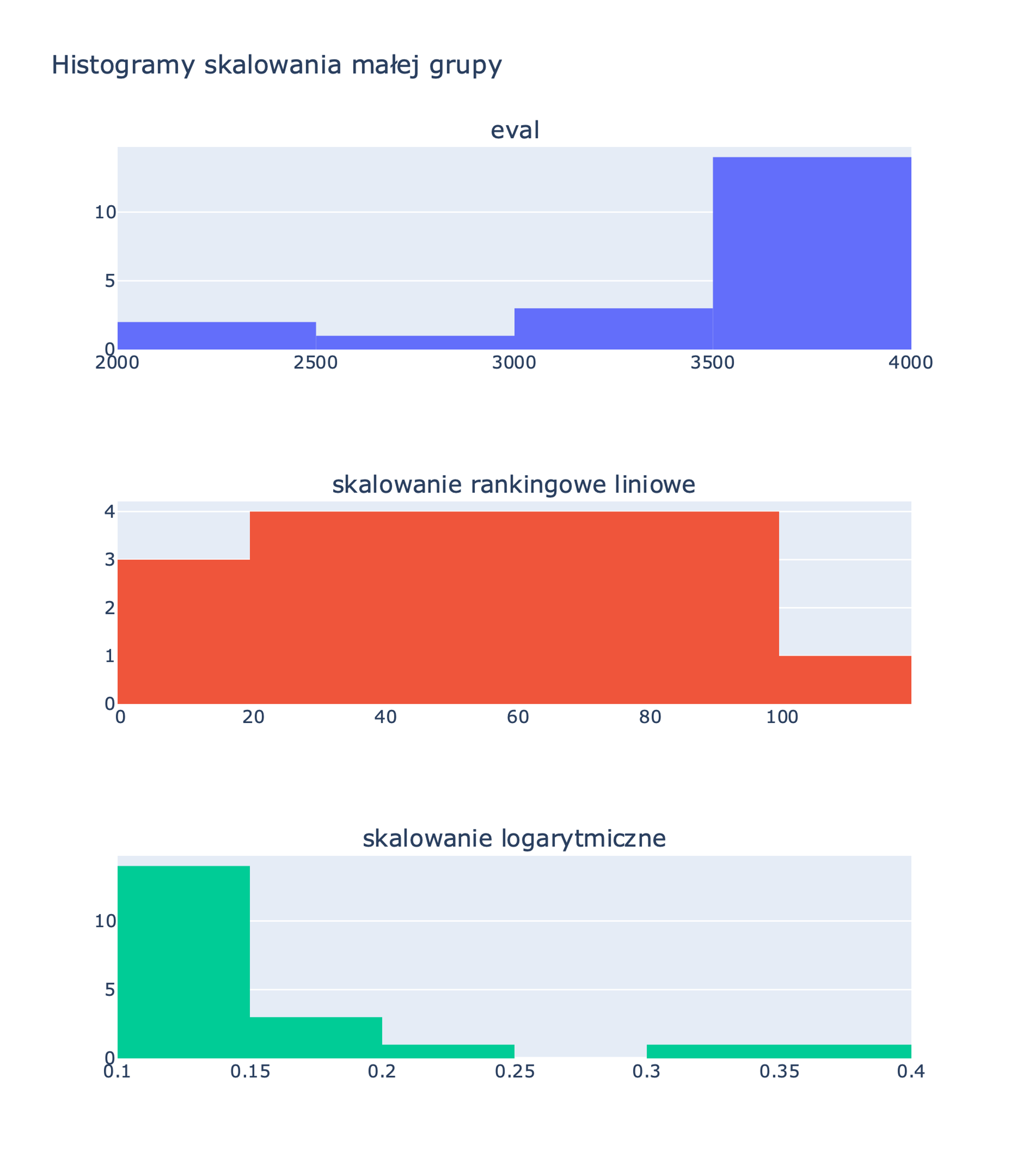
A screenshot of a cell phone

Description automatically generated

Rysunek 15 - Parametry statystyczne skalowania małej grupy dla siodła Rosenbrocka



Rysunek 16 - Wykres skrzypcowy małej grupy dla siodła Rosenbrocka



Rysunek 17 - Histogramy małej grupy dla siodła Rosenbrocka

## Duży zbiór danych

Pierwszy badany zbiór składał się z 100 elementów. Poniżej przedstawione są wyliczone dla niego wartości oraz analiza tych danych.

A screenshot of a cell phone

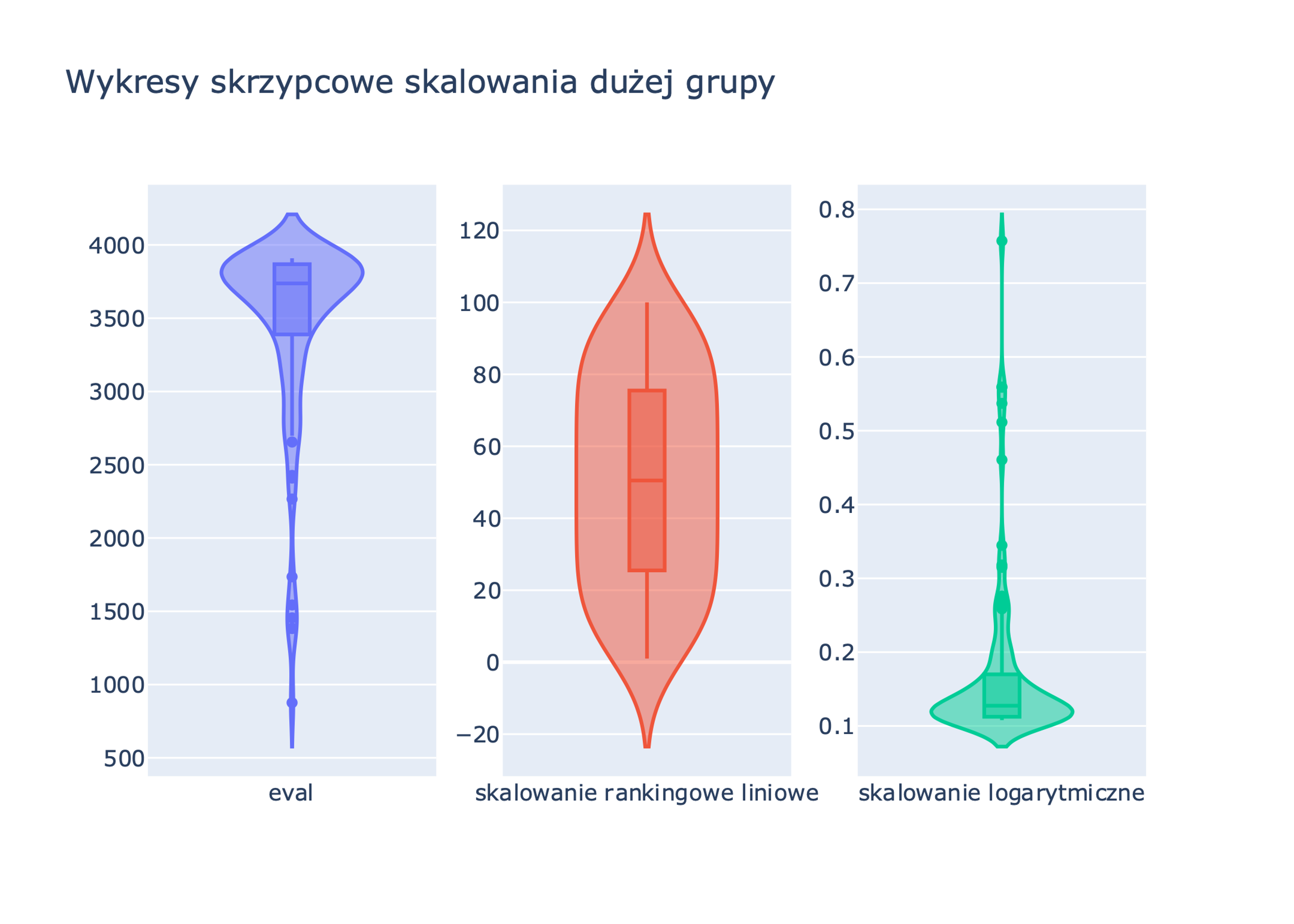
Description automatically generated

Rysunek 18 - Przykładowe wartości dużej grupy dla siodła Rosenbrocka

A screenshot of a cell phone

Description automatically generated

Rysunek 19 - Parametry statystyczne skalowania dużej grupy dla siodła Rosenbrocka



Rysunek 20 - Wykresy skrzypcowe dużej grupy dla siodła Rosenbrocka



Rysunek 21 - Histogramy dużej grupy dla siodła Rosenbrocka

## Wnioski

Wykresy skrzypcowe oraz histogramy funkcji ewaluacji i skalowania logarytmicznego są symetryczne odbite względem osi y. Świadczy to o odwróceniu kolejności danych poprzez skalowanie liniowe.

Wykres skrzypcowy skalowania rankingowego liniowego jest równo podzielony na cztery kwartyle. Świadczy to o równym rozkładzie punktów. Potwierdza to również praktycznie idealnie płaski histogram tego skalowania.

Bardzo mały odchyleniu standardowym oraz skali wykresu skrzypcowego można zaobserwować, że skalowanie logarytmiczne mocno zbliża punkty do siebie oraz do początku układu współrzędnych. Widać to również po zmianie wartości minimalnej i maksymalnej przeskalowanych wartości.

Na wykresie pudełkowym wartości funkcji ewaluacji można zaobserwować punkty bardzo oddalone od reszty. Jest to szczególnie widoczne na wykresie dużej grupy testowej. Dane te świadczą o gwałtownym spadku wartości funkcji w pobliżu jej ekstremum. Po rozkładzie kwartyli oraz jądrowego estymatora gęstości można zaobserwować, że obszar o znacznym spadku wartości funkcji jest stosunkowo niewielki.