

Wspomaganie Decyzji

Laboratorium 5

Podstawy programowania w systemie SAS – SAS1

1. Utwórz bibliotekę o nazwie Lab5.

```
libname Lab5 '/folders/myfolders/Lab5';
```

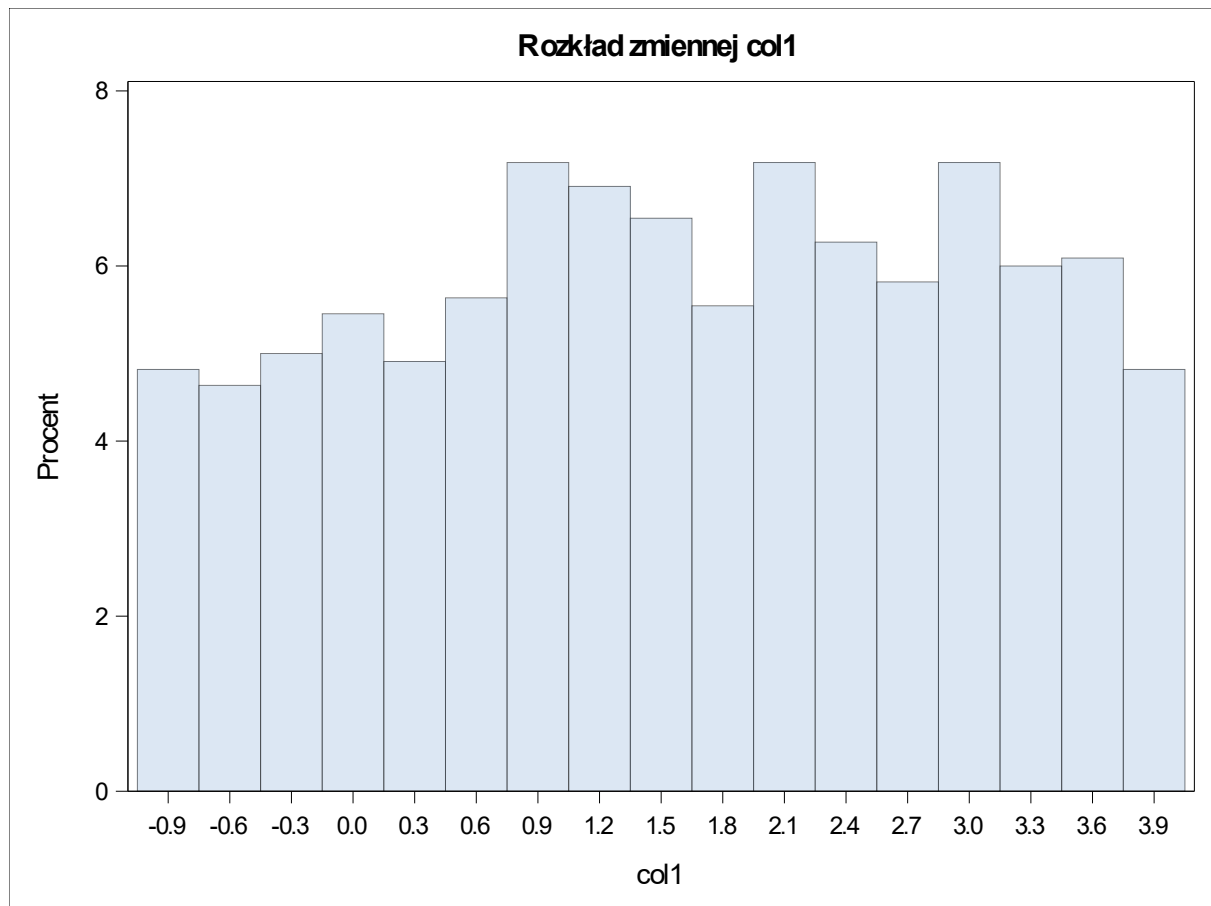
2. Wczytaj tekstowy plik_1 do pliku sas o nazwie dane_lab5 w bibliotece Lab5.

```
data Lab5.dane_lab5;  
    infile '/folders/myfolders/Lab5/dane_4.txt' dlm = ";";  
    input id text $ col1 col2 col3;  
run;
```

3. Narysuj histogramy dla zmiennych numerycznych.

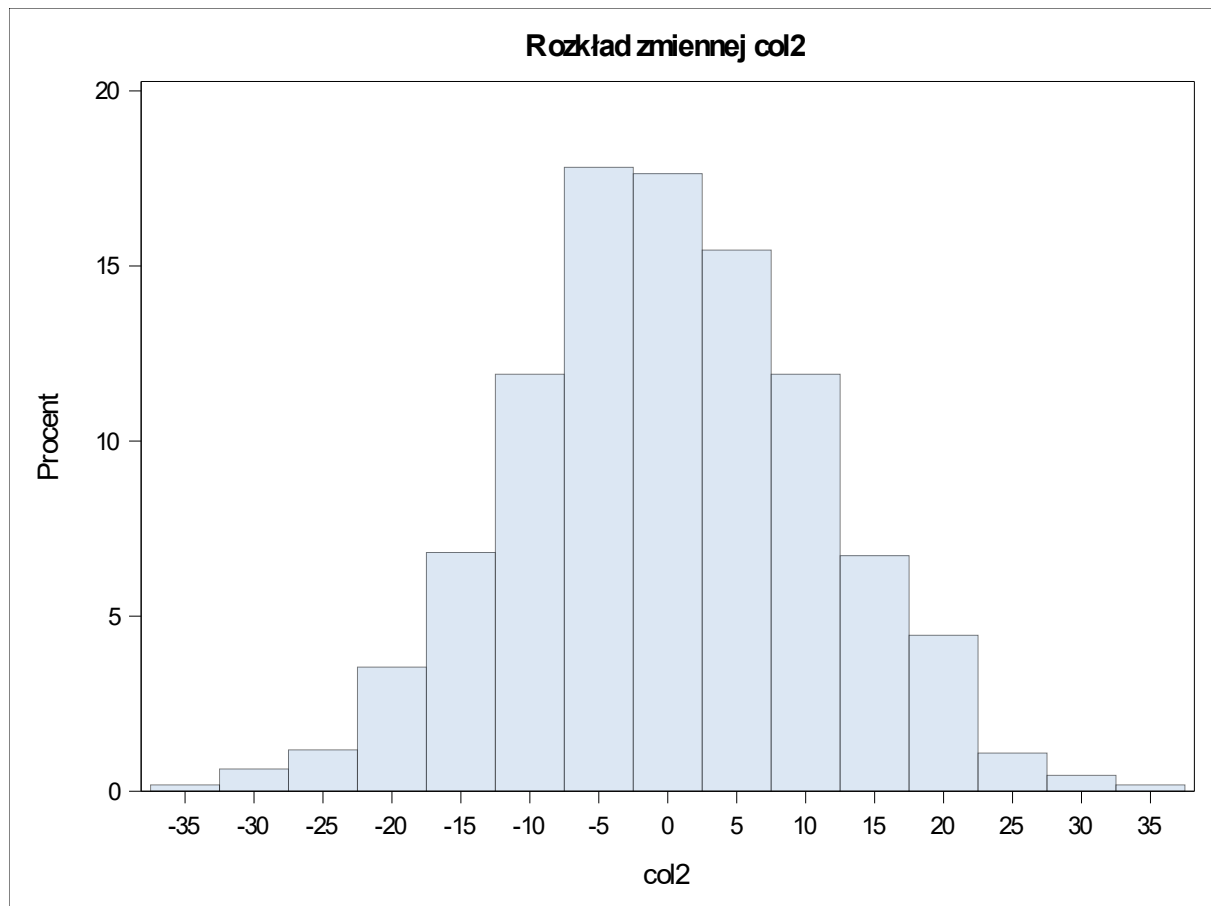
Kolumna danych numerycznych col1:

```
/* Histogram dla kolumny danych col1 */  
PROC UNIVARIATE data = Lab5.dane_lab5;  
    title 'Histogram - col1';  
    VAR col1;  
    HISTOGRAM col1;  
run;
```



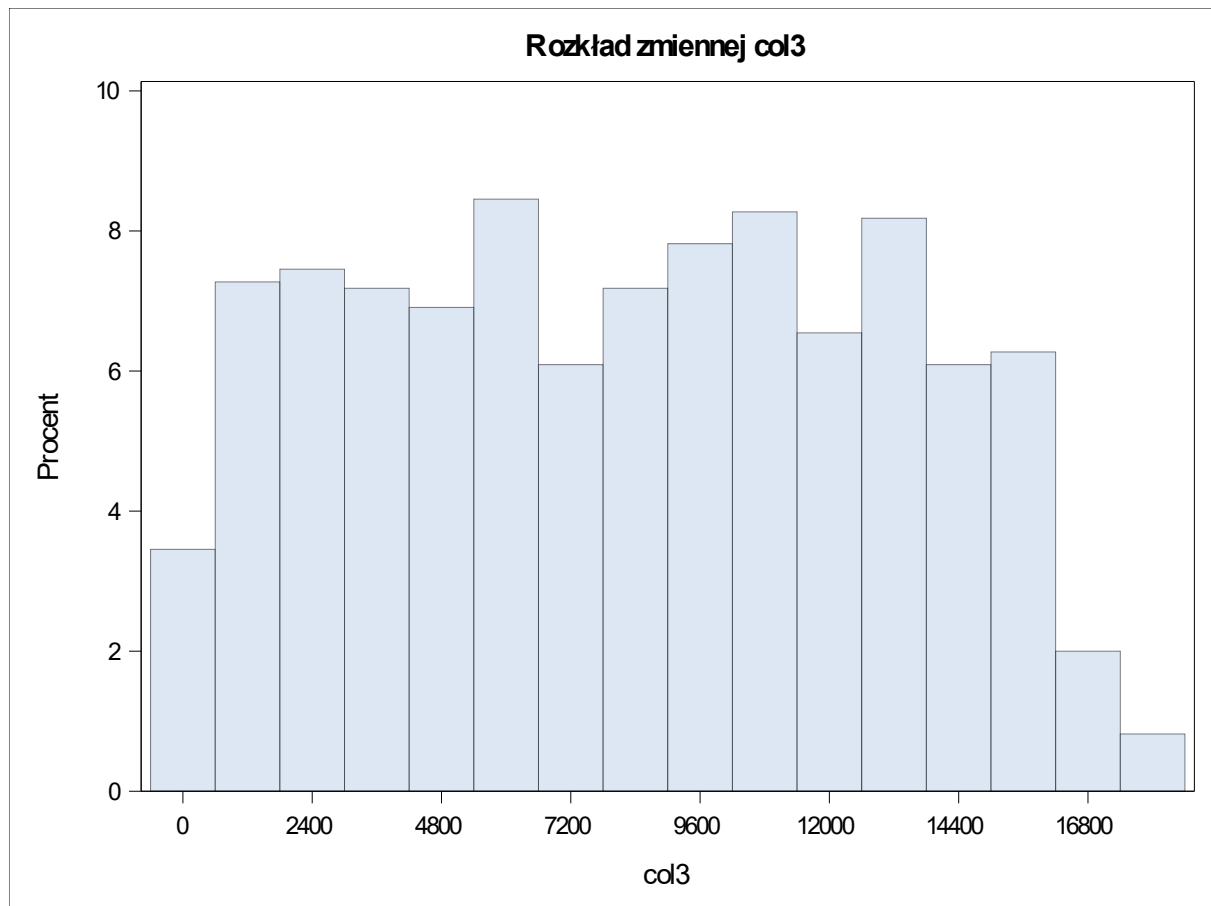
Kolumna danych numerycznych col2:

```
/* Histogram dla kolumny danych col2 */  
PROC UNIVARIATE data = Lab5.dane_lab5;  
    title 'Histogram - col2';  
    VAR col2;  
    HISTOGRAM col2;  
run;
```



Kolumna danych numerycznych col3:

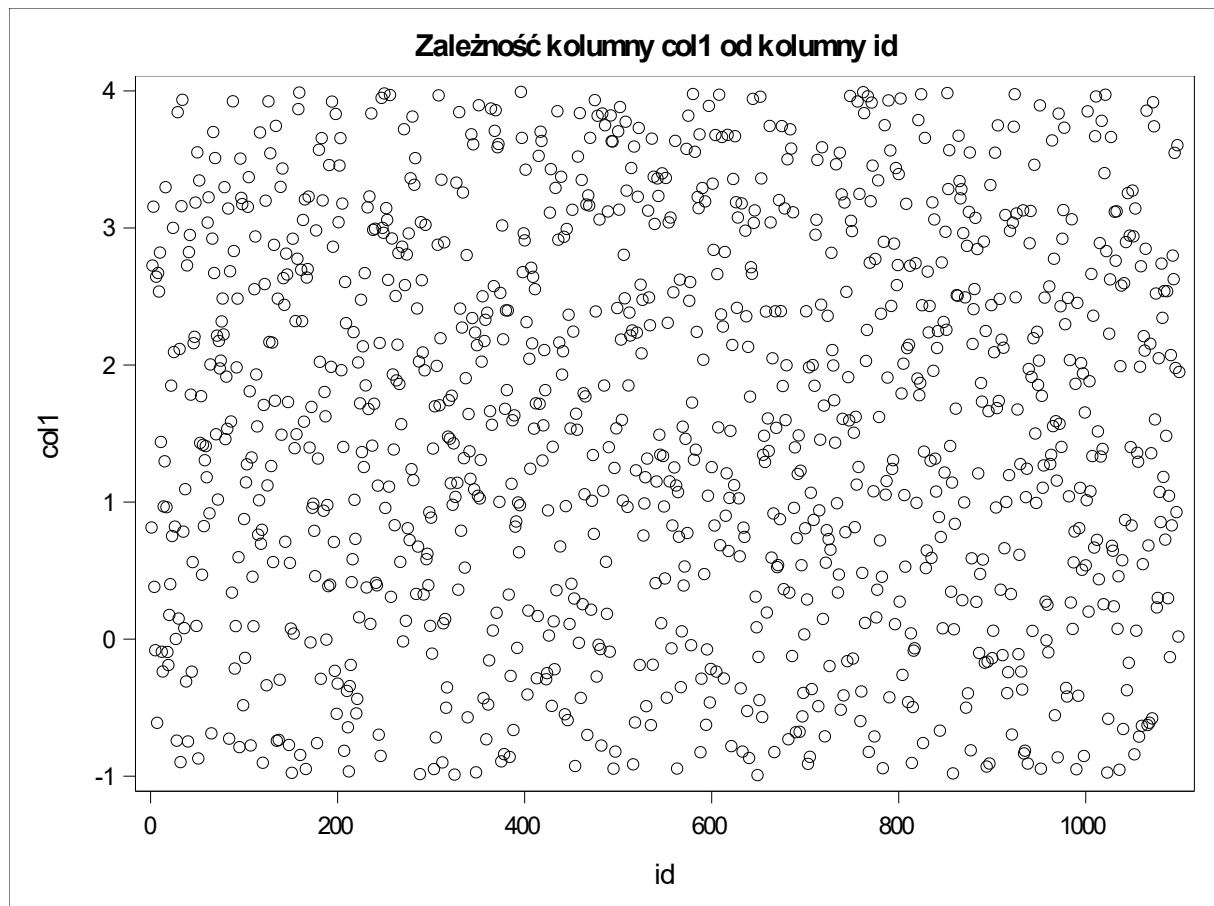
```
/* Histogram dla kolumny danych col3 */  
PROC UNIVARIATE data = Lab5.dane_lab5;  
    title 'Histogram - col3';  
    VAR col3;  
    HISTOGRAM col3;  
run;
```



4. Narysuj wykresy zależności zmiennych numerycznych od id.

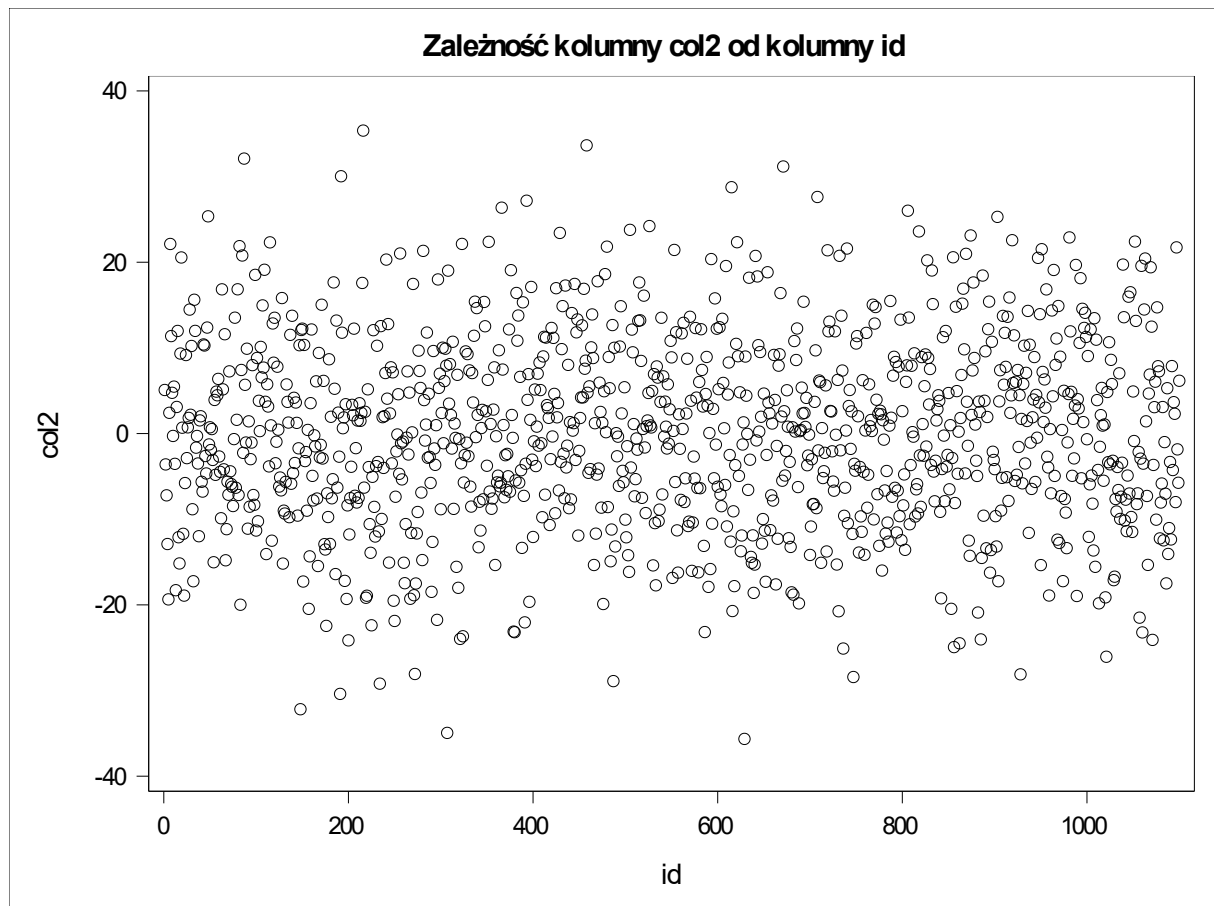
Kolumna danych numerycznych col1:

```
/* col1 = f(id) */  
proc sgplot data = Lab5.dane_lab5;  
    title 'Zależność kolumny col1 od kolumny id';  
    scatter x = id y = col1; /* wykres punktowy */  
run;
```



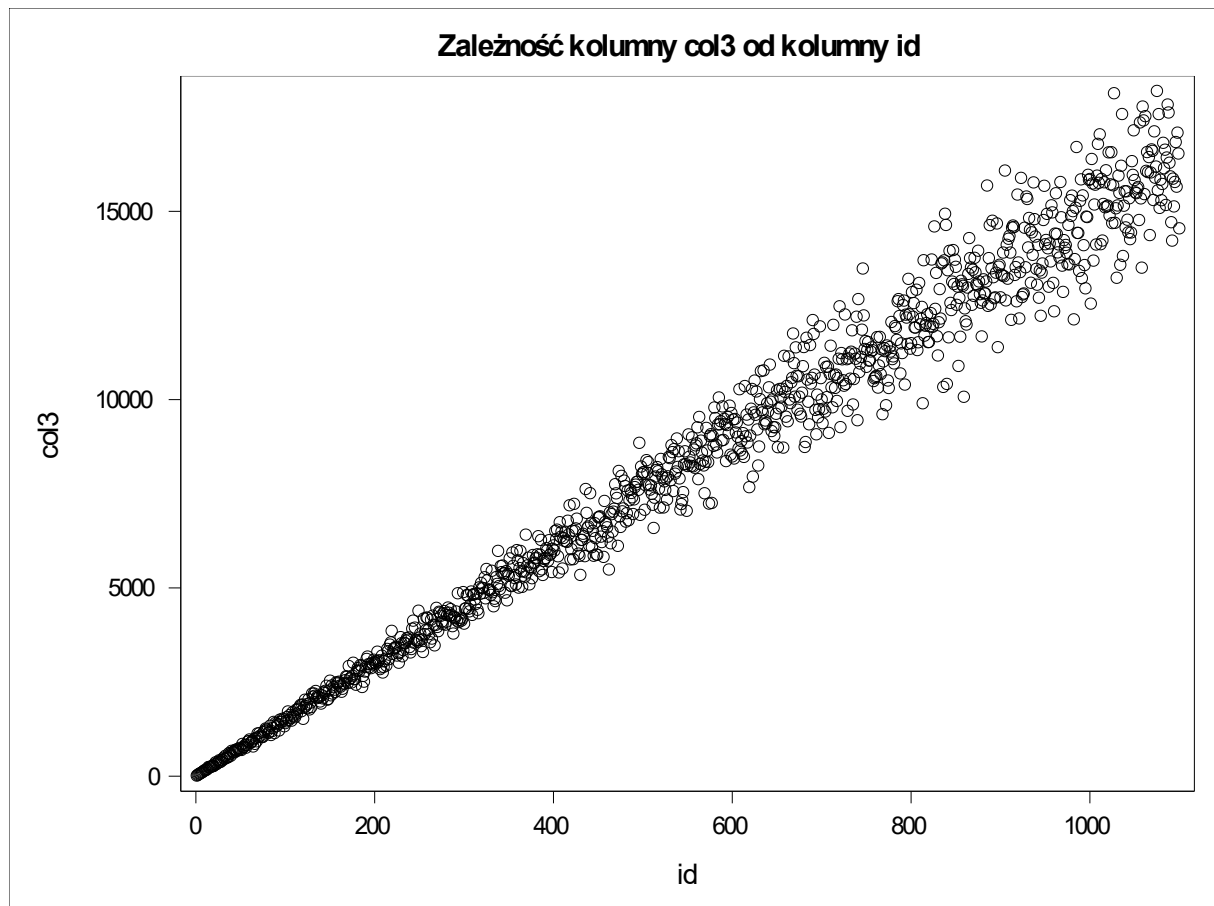
Kolumna danych numerycznych col2:

```
/* col2 = f(id) */  
proc sgplot data = Lab5.dane_lab5;  
    title 'Zależność kolumny col2 od kolumny id';  
    scatter x = id y = col2; /* wykres punktowy */  
run;
```



Kolumna danych numerycznych col3:

```
/* col3 = f(id) */  
proc sgplot data = Lab5.dane_lab5;  
    title 'Zależność kolumny col3 od kolumny id';  
    scatter x = id y = col3; /* wykres punktowy */  
run;
```



5. Policz za pomocą procedury MEANS statystyki: N, MAX, MIN, NMISS, MEAN.

```
PROC MEANS DATA = Lab5.dane_lab5 N MAX MIN NMISS MEAN;  
title 'Parametry kolumn col1, col2, col3';  
VAR col1 col2 col3;  
RUN;
```

Zmienna	N	Maksimum	Minimum	N braków	Średnia
col1	1100	3.9929553	-0.9920752	0	1.5910807
col2	1100	35.3683514	-35.6461345	0	0.0732446
col3	1100	18198.25	16.4508137	0	8255.33

6. Policz statystyki oddzielnie dla id parzystych i nieparzystych z wykorzystaniem CLASS i BY.

```
/* Wykorzystanie CLASS */
proc sort data = Lab5.dane_lab5;
by text;
run;

PROC MEANS DATA = Lab5.dane_lab5;
title 'Parametry kolumn col1, col2, col3, w zależności od text';
VAR col1 col2 col3;
CLASS text; /* dane w 1 tabeli */
RUN;
```

text	N obs.	Zmienna	N	Średnia	Odch. std.	Minimum	Maksimum
s4_niep	550	col1	550	1.6165933	1.3787437	-0.9920752	3.9880219
		col2	550	0.3862441	11.1956235	-35.6461345	32.0996697
		col3	550	8278.88	4803.49	16.4508137	18198.25
s4_parz	550	col1	550	1.5655681	1.4395945	-0.9847683	3.9929553
		col2	550	-0.2397549	10.8808746	-32.1838829	35.3683514
		col3	550	8231.77	4770.23	31.1938689	17625.39


```

/* Wykorzystanie BY */
proc sort data = Lab5.dane_lab5;
by text;
run;

PROC MEANS DATA = Lab5.dane_lab5;
title 'Parametry kolumn col1, col2, col3, w zależności od text';
VAR col1 col2 col3;
BY text; /* dane w oddzielnych tabelach */
RUN;

```

text=s4_niep

Zmienna	N	Średnia	Odch. std.	Minimum	Maksimum
col1	550	1.6165933	1.3787437	-0.9920752	3.9880219
col2	550	0.3862441	11.1956235	-35.6461345	32.0996697
col3	550	8278.88	4803.49	16.4508137	18198.25

text=s4_parz

Zmienna	N	Średnia	Odch. std.	Minimum	Maksimum
col1	550	1.5655681	1.4395945	-0.9847683	3.9929553
col2	550	-0.2397549	10.8808746	-32.1838829	35.3683514
col3	550	8231.77	4770.23	31.1938689	17625.39

7. Znajdź parametry **a** i **b** modelu regresji $y = a + b \cdot id$ dla id parzystych.

UWAGA. Założyłem, że to zadanie wykonam dla każdej z kolumn danych numerycznych col1, col2, col3.

Kolumna danych numerycznych col1:

```
/* col1 = f(id) */
PROC REG data = Lab5.dane_lab5;
title 'Zależność kolumny col1 od id - regresja liniowa dla id
parzystych';
MODEL col1 = id;
where mod(id,2) = 0;
RUN; /* y= a + b*id dla b_col1_parz= -0.00024570, a_col1_parz =
1.70095 */
```

Oceny parametrów					
Zmienna	DF	Ocena parametru	Błąd standardowy	Wartość t	Pr. > t
Intercept	1	1.70095	0.12287	13.84	<.0001
id	1	-0.00024570	0.00019320	-1.27	0.2040

Zatem jeśli $y = a + b \cdot id$, to $a = 1.70095$, $b = -0.00024570$.

Kolumna danych numerycznych col2:

```
/* col2 = f(id) */
PROC REG data = Lab5.dane_lab5;
title 'Zależność kolumny col2 od id - regresja liniowa dla id
parzystych';
MODEL col2 = id;
where mod(id,2) = 0;
RUN; /* y= a + b*id dla b_col2_parz= 0.00063185, a_col2_parz =
-0.58791 */
```

Oceny parametrów					
Zmienna	DF	Ocena parametru	Błąd standardowy	Wartość t	Pr. > t
Intercept	1	-0.58791	0.92988	-0.63	0.5275
id	1	0.00063185	0.00146	0.43	0.6658

Zatem jeśli $y = a + b \cdot id$, to $a = -0.58791$, $b = 0.00063185$.

Kolumna danych numerycznych col3:

```
/* col3 = f(id) */
PROC REG data = Lab5.dane_lab5;
title 'Zależność kolumny col3 od id - regresja liniowa dla id
parzystych';
MODEL col3 = id;
where mod(id,2) = 0;
RUN; /* y= a + b*id dla b_col3_parz= 14.86533, a_col3_parz =
40.97727 */
```

Oceny parametrów					
Zmienna	DF	Ocena parametru	Błąd standardowy	Wartość t	Pr. > t
Intercept	1	40.97727	56.22020	0.73	0.4664
id	1	14.86533	0.08840	168.15	<.0001

Zatem, jeśli $y = a + b \cdot id$, to $a = 40.97727$, $b = 14.86533$.

8. Dodaj kolumnę o nazwie ye zawierającą obliczona wartość z modelu regresji dla id nieparzystych.

UWAGA. Założyłem, że to zadanie wykonam dla każdej z kolumn danych numerycznych col1, col2, col3.

Kolumna danych numerycznych col1:

```
/* col1 */
/* col1 = f(id) */
PROC REG data = Lab5.dane_lab5;
title 'Zależność kolumny col1 od id - regresja liniowa dla id
nieparzystych';
MODEL col1 = id;
where mod(id,2) = 1;
RUN; /* y= a + b*id dla b_col1_nparz = 0.00014486      ,
a_col1_nparz = 1.53692 */

/* wstawienie kolumny ye_col1_nparz */
data Lab5.dane_lab5;
set Lab5.dane_lab5;
b_col1_nparz = 0.00014486;
a_col1_nparz = 1.53692;
ye_col1_nparz = a_col1_nparz + b_col1_nparz * id;
where mod(id,2) = 1;
run;
```

Parametry a, b regresji $y = a + b \cdot id$ wynoszą $a = 1.53692$, $b = 0.00014486$ (tabela poniżej).

Oceny parametrów					
Zmienna	DF	Ocena parametru	Błąd standardowy	Wartość t	Pr. > t
Intercept	1	1.53692	0.11762	13.07	<.0001
id	1	0.00014486	0.00018521	0.78	0.4345

Kolumna danych numerycznych col2:

```
/* col2 */
/* col2 = f(id) */
PROC REG data = Lab5.dane_lab5;
title 'Zależność kolumny col2 od id - regresja liniowa dla id
nieparzystych';
MODEL col2 = id;
where mod(id,2) = 1;
RUN; /* y= a + b*id dla b_col2_nparz = 0.00062574      ,
a_col2_nparz = 0.04208 */

/* wstawienie kolumny ye_col2_nparz */
data Lab5.dane_lab5;
```

```

set Lab5.dane_lab5;
b_col2_nparz = 0.00062574;
a_col2_nparz = 0.04208;
ye_col2_nparz = a_col2_nparz + b_col2_nparz * id;
where mod(id,2) = 1;
run;

```

Parametry a, b regresji $y = a + b \cdot id$ wynoszą $a = 0.04208$, $b = 0.00062574$ (tabela poniżej).

Oceny parametrów					
Zmienna	DF	Ocena parametru	Błąd standardowy	Wartość t	Pr. > t
Intercept	1	0.04208	0.95549	0.04	0.9649
id	1	0.00062574	0.00150	0.42	0.6776

Kolumna danych numerycznych col3:

```

/* col3 */
/* col3 = f(id) */
PROC REG data = Lab5.dane_lab5;
title 'Zależność kolumny col3 od id - regresja liniowa dla id
nieparzystych';
MODEL col3 = id;
where mod(id,2) = 1;
RUN; /* y=a*id+b dla a_col3_nparz = 14.95392      , b_col3_nparz =
54.22421 */

/* wstawienie kolumny ye_col3_nparz */
data Lab5.dane_lab5;
set Lab5.dane_lab5;
a_col3_nparz = 14.95392;
b_col3_nparz = 54.22421;
ye_col3_nparz = a_col3_nparz * id + b_col3_nparz;
where mod(id,2) = 1;
run;

```

Parametry a, b regresji $y = a + b \cdot id$ wynoszą $a = 54.22421$, $b = 14.95392$ (tabela poniżej).

Oceny parametrów					
Zmienna	DF	Ocena parametru	Błąd standardowy	Wartość t	Pr. > t
Intercept	1	54.22421	59.39561	0.91	0.3617
id	1	14.95392	0.09352	159.89	<.0001

9. Narysuj zależność y i ye od id nieparzystych dla 10, 100 i wszystkich obserwacji (nieparzystych).

UWAGA. Założyłem, że to zadanie wykonam dla każdej z kolumn danych numerycznych col1, col2, col3.

Kolumna danych col1:

```
/* col1 */
/* wszystkie punkty o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
    title 'col1(id) oraz ye_col1_nparz(id) - wszystkie id
nieparzyste'; /* tytuł wykresu */
    scatter x = id y = col1; /* wykres punktowy */
    scatter x = id y = ye_col1_nparz; /* wykres punktowy */
    where mod(id,2)=1;
run;

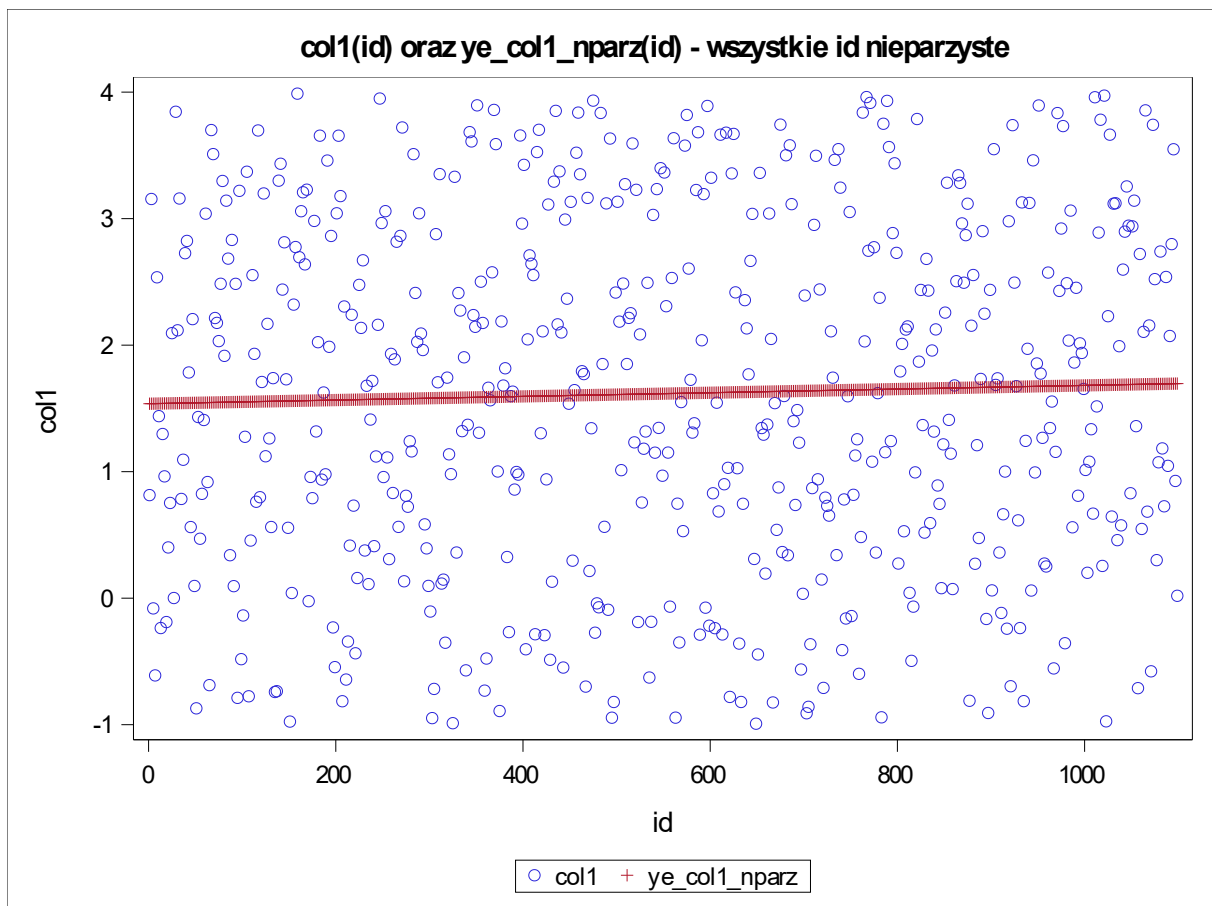
/* 10 punktów o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
```

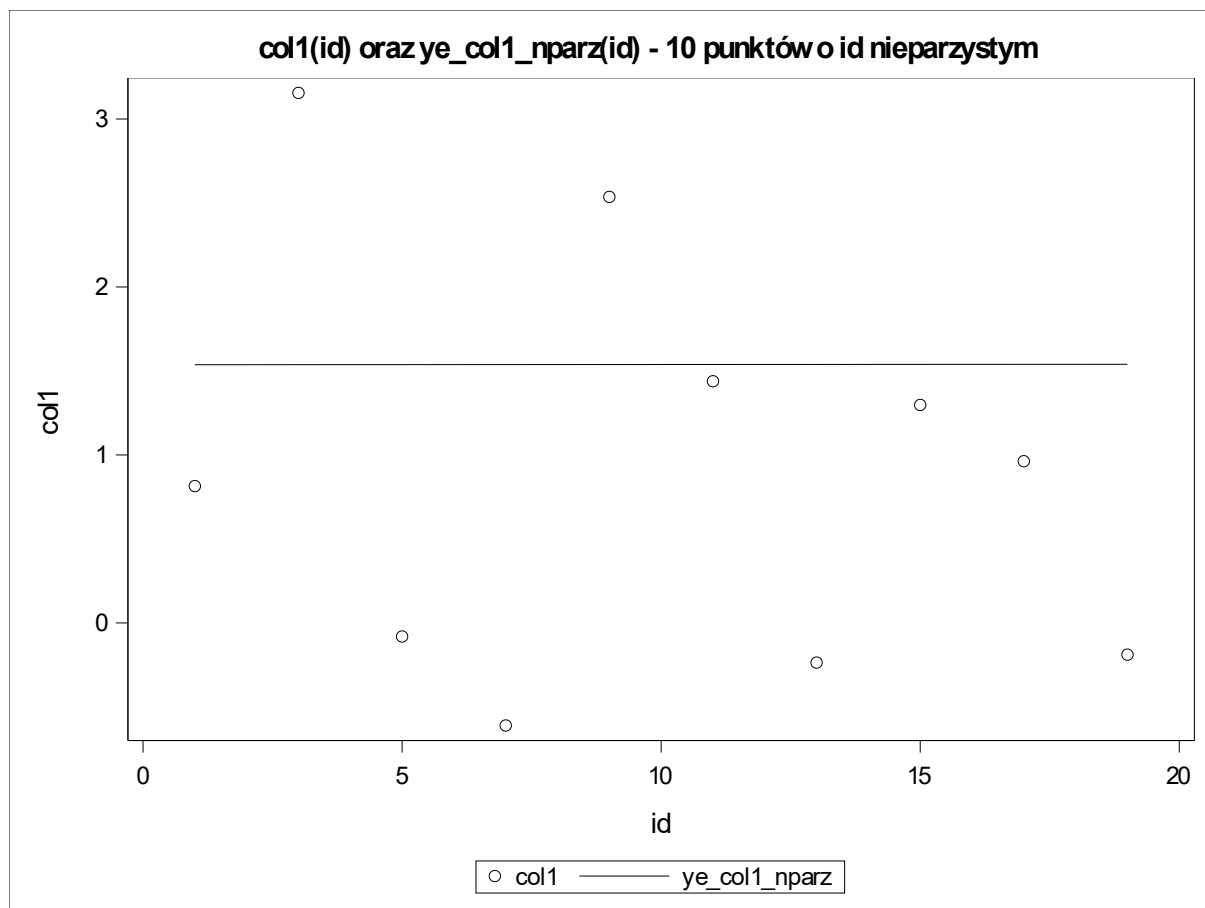
```

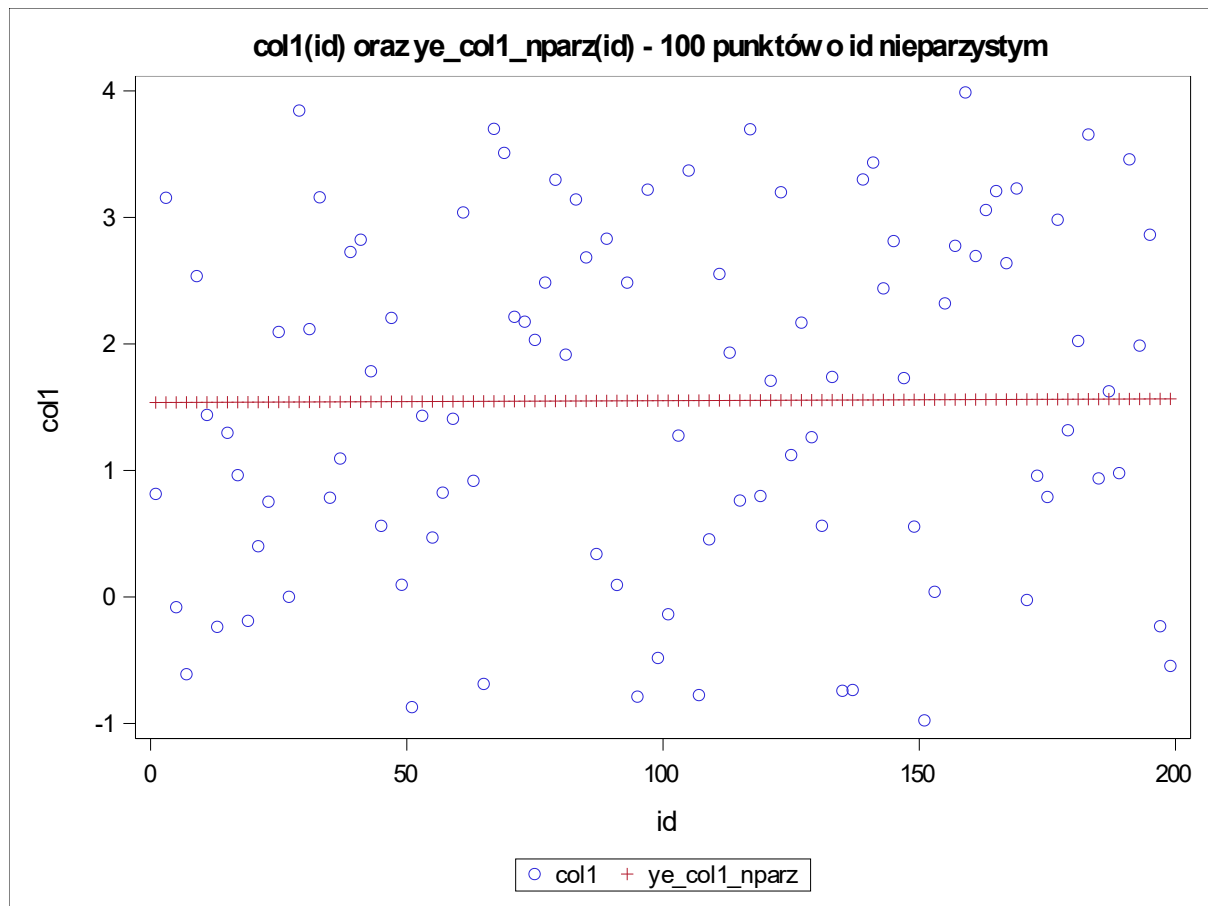
        title 'col1(id) oraz ye_col1_nparz(id) - 10 punktów o id
nieparzystym'; /* tytuł wykresu */
        scatter x = id y = col1; /* wykres punktowy */
        series x = id y = ye_col1_nparz; /* wykres liniowy */
        where mod(id,2)=1 and id <= 19;
run;

/* 100 punktów o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
        title 'col1(id) oraz ye_col1_nparz(id) - 100 punktów o id
nieparzystym'; /* tytuł wykresu */
        scatter x = id y = col1; /* wykres punktowy */
        scatter x = id y = ye_col1_nparz; /* wykres punktowy */
        where mod(id,2) = 1 and id <= 199;
run;

```







Kolumna danych col2:

```
/* col2 */
/* wszystkie punkty o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
    title 'col2(id) oraz ye_col2_nparz(id) - wszystkie id
nieparzyste'; /* tytuł wykresu */
    scatter x = id y = col2; /* wykres punktowy */
    scatter x = id y = ye_col2_nparz; /* wykres punktowy */
    where mod(id,2)=1;
run;

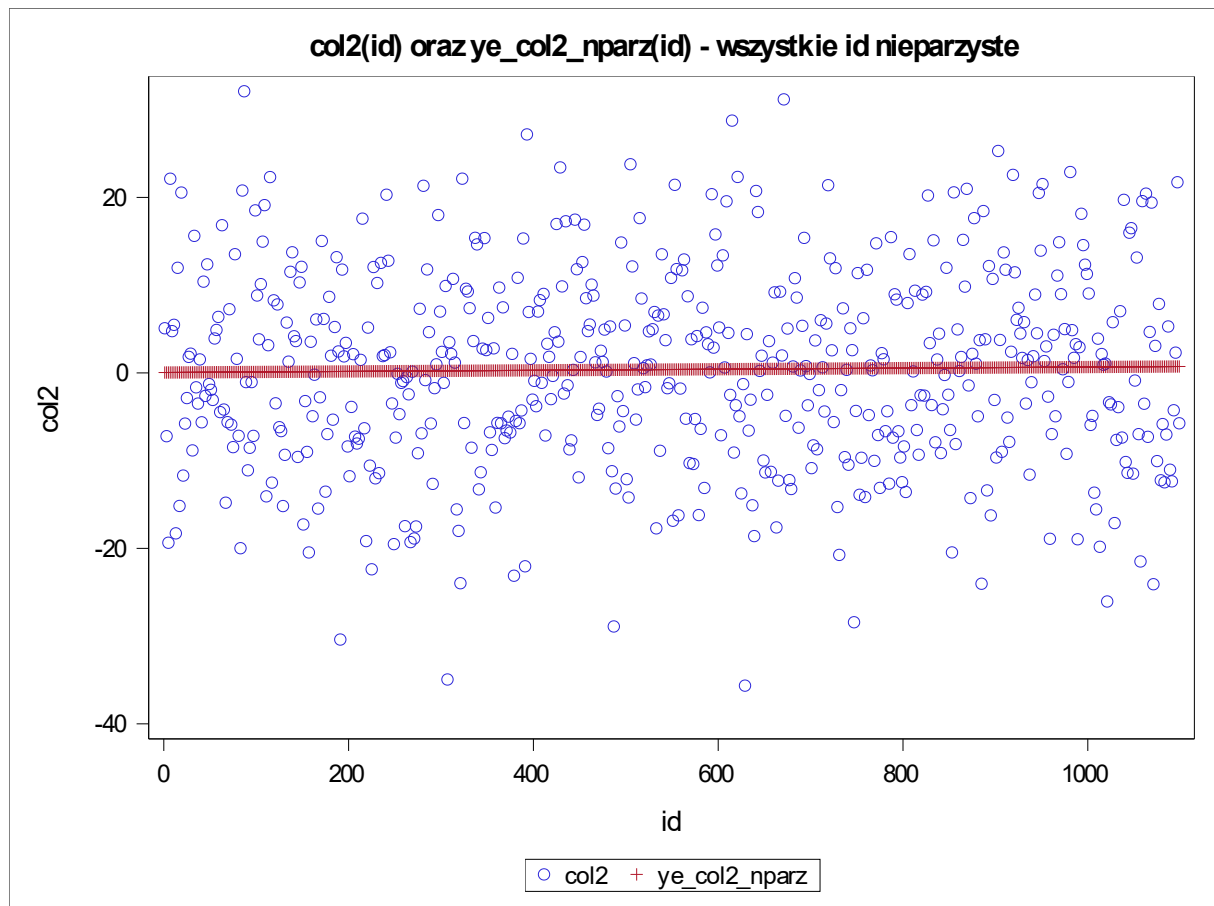
/* 10 punktów o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
    title 'col2(id) oraz ye_col2_nparz(id) - 10 punktów o id
nieparzystym'; /* tytuł wykresu */
    scatter x = id y = col2; /* wykres punktowy */
    series x = id y = ye_col2_nparz; /* wykres liniowy */
    where mod(id,2)=1 and id <= 19;
run;

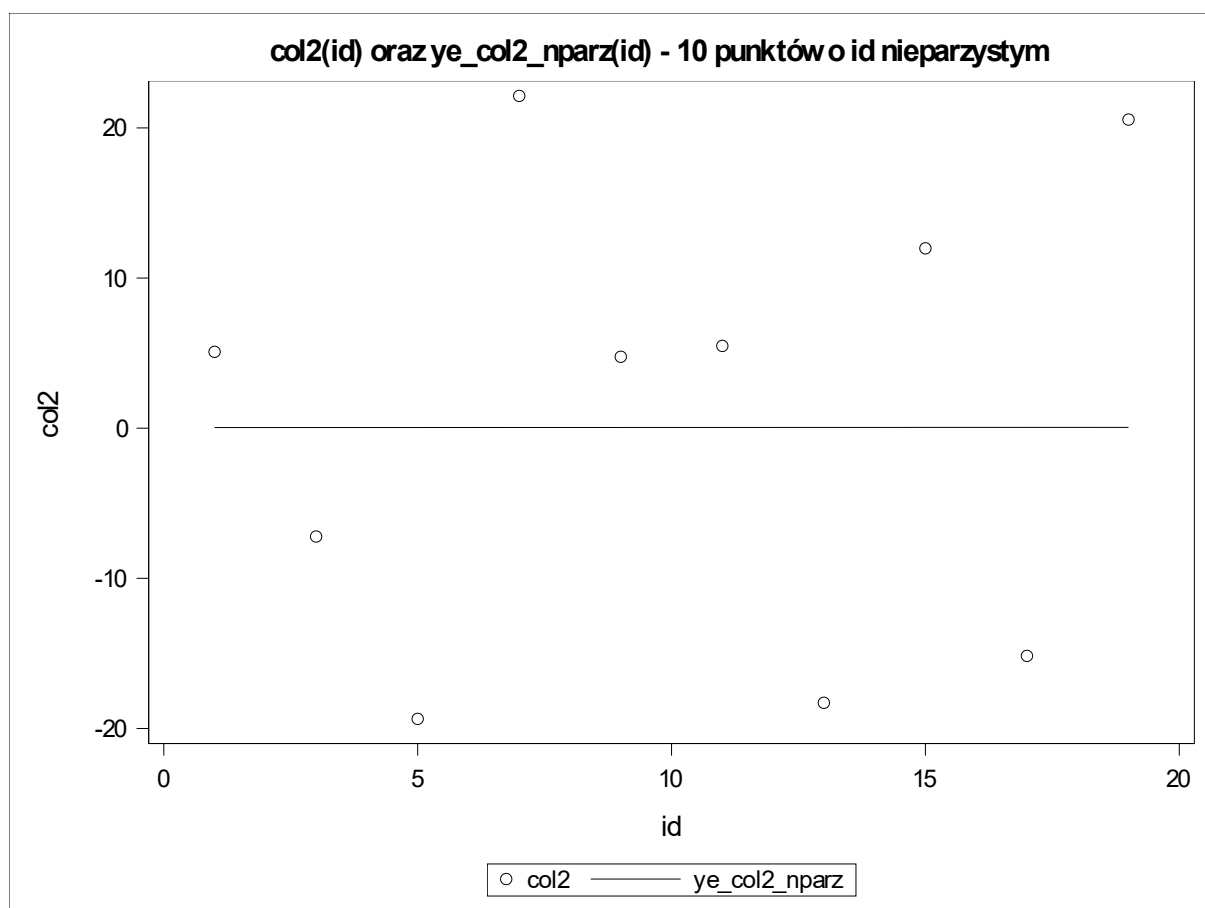
/* 100 punktów o id nieparzystym */
```

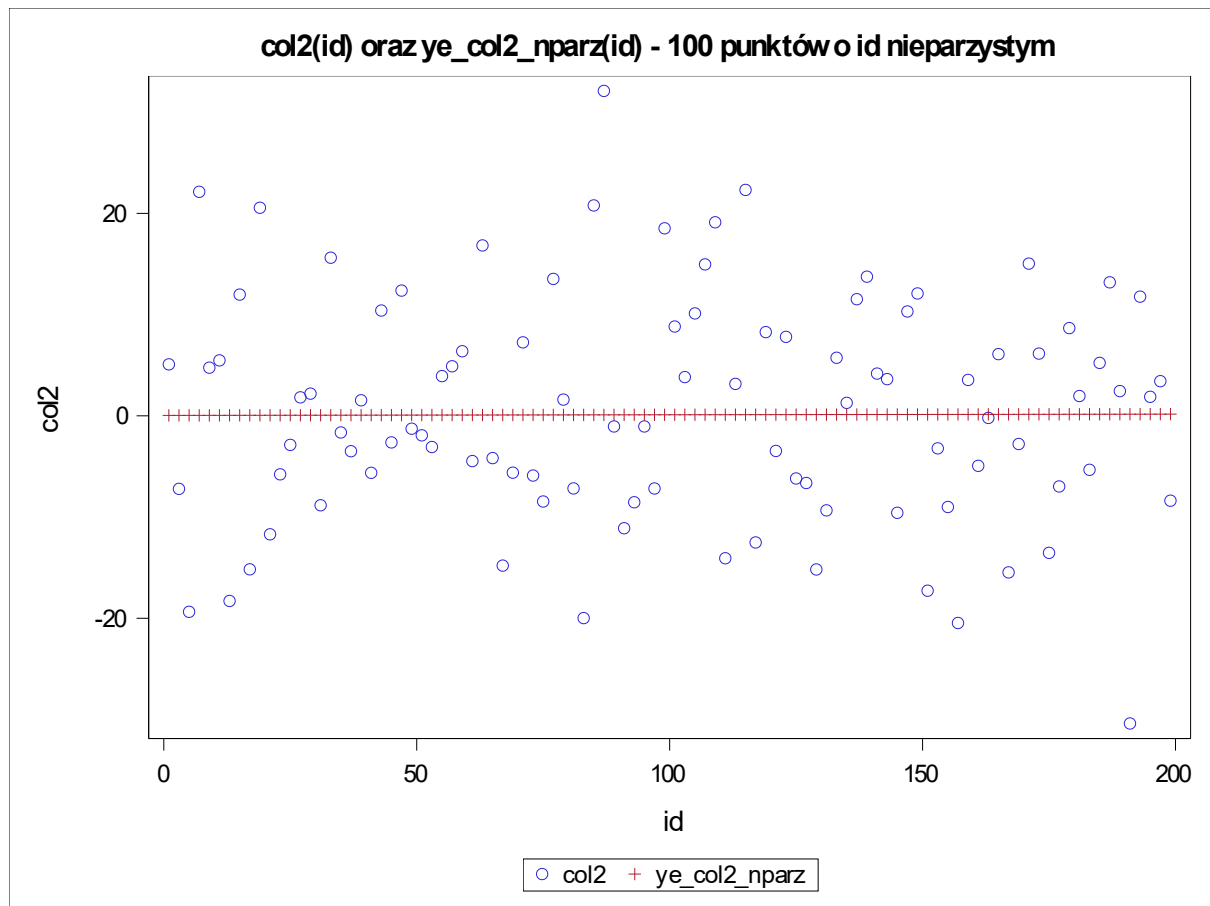
```

proc sgplot data = Lab5.dane_lab5;
    title 'col2(id) oraz ye_col2_nparz(id) - 100 punktów o id
nieparzystym'; /* tytuł wykresu */
    scatter x = id y = col2; /* wykres punktowy */
    scatter x = id y = ye_col2_nparz; /* wykres punktowy */
    where mod(id,2) = 1 and id <= 199;
run;

```







Kolumna danych col3:

```
/* col3 */
/* wszystkie punkty o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
    title 'col3(id) oraz ye_col3_nparz(id) - wszystkie id
nieparzyste'; /* tytuł wykresu */
    scatter x = id y = col3; /* wykres punktowy */
    scatter x = id y = ye_col3_nparz; /* wykres punktowy */
    where mod(id,2)=1;
run;

/* 10 punktów o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
    title 'col3(id) oraz ye_col3_nparz(id) - 10 punktów o id
nieparzystym'; /* tytuł wykresu */
    scatter x = id y = col3; /* wykres punktowy */
    series x = id y = ye_col3_nparz; /* wykres liniowy */
    where mod(id,2)=1 and id <= 19;
run;
```

```

/* 100 punktów o id nieparzystym */
proc sgplot data = Lab5.dane_lab5;
    title 'col3(id) oraz ye_col3_nparz(id) - 100 punktów o id
nieparzystym'; /* tytuł wykresu */
    scatter x = id y = col3; /* wykres punktowy */
    scatter x = id y = ye_col3_nparz; /* wykres punktowy */
    where mod(id,2) = 1 and id <= 199;
run;

```

