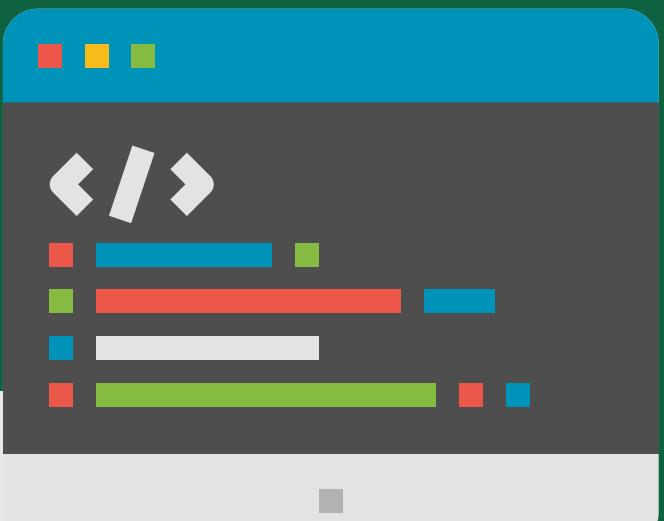


# Modele regresyjne



**Autorzy:**  
**Weronika Forszpaniak**  
**Kiryl Radziul**



# WPROWADZENIE

Analiza modelu

Zbiory danych

Wykresy

## AGENDA

- 1. Opis zbiorów danych**
- 2. Model podstawowy ze zmiennymi jakościowymi**
- 3. Model podstawowy + współliniowość**
- 4. Model z moderacją**

Podsumowanie

Uzasadnienie merytoryczne



# OPIS ZBIORÓW

Rows: 176

Columns: 4

\$ weight <dbl> 240, 250, 255, 260, 262, 258, 266, 266, 265,  
\$ Time <dbl> 1, 8, 15, 22, 29, 36, 43, 44, 50, 57, 64, 1,  
\$ Rat <fct> A, A, A, A, A, A, A, A, A, B, B, B, B,  
\$ Diet <fct> a, a,

## \* ZBIÓR “BODYWEIGHT”

Opisuje dane dotyczące masy ciała (w gramach) szczurów mierzone przez 64 dni. Istnieją trzy grupy szczurów, każda na innej diecie (a, b i c). Zbiór ten zawiera 176 wierszy i 4 kolumny.



# DANYCH

## \* ZBIÓR “FAT”

Pomiar procentu tkanki tłuszczowej dla 252 mężczyzn za pomocą techniki ważenia pod wodą, a także pomiar obwodu różnych części ciała. Zbiór zawiera 252 wiersze i 18 kolumn.

Rows: 252

Columns: 18

\$ brozek <dbl> 12.6, 6.9, 24.6, 10.9, 27.8,  
\$ siri <dbl> 12.3, 6.1, 25.3, 10.4, 28.7,  
\$ density <dbl> 1.0708, 1.0853, 1.0414, 1.0751  
\$ age <int> 23, 22, 22, 26, 24, 24, 26,  
\$ weight <dbl> 154.25, 173.25, 154.00, 184.75  
\$ height <dbl> 67.75, 72.25, 66.25, 72.25,  
\$ adipos <dbl> 23.7, 23.4, 24.7, 24.9, 25.6,  
\$ free <dbl> 134.9, 161.3, 116.0, 164.7,  
\$ neck <dbl> 36.2, 38.5, 34.0, 37.4, 34.4,  
\$ chest <dbl> 93.1, 93.6, 95.8, 101.8, 97.3,  
\$ abdom <dbl> 85.2, 83.0, 87.9, 86.4, 100.0,  
\$ hip <dbl> 94.5, 98.7, 99.2, 101.2, 101.9  
\$ thigh <dbl> 59.0, 58.7, 59.6, 60.1, 63.2,  
\$ knee <dbl> 37.3, 37.3, 38.9, 37.3, 42.2,  
\$ ankle <dbl> 21.9, 23.4, 24.0, 22.8, 24.0,  
\$ biceps <dbl> 32.0, 30.5, 28.8, 32.4, 32.2,  
\$ forearm <dbl> 27.4, 28.9, 25.2, 29.4, 27.7,  
\$ wrist <dbl> 17.1, 18.2, 16.6, 18.2, 17.7,

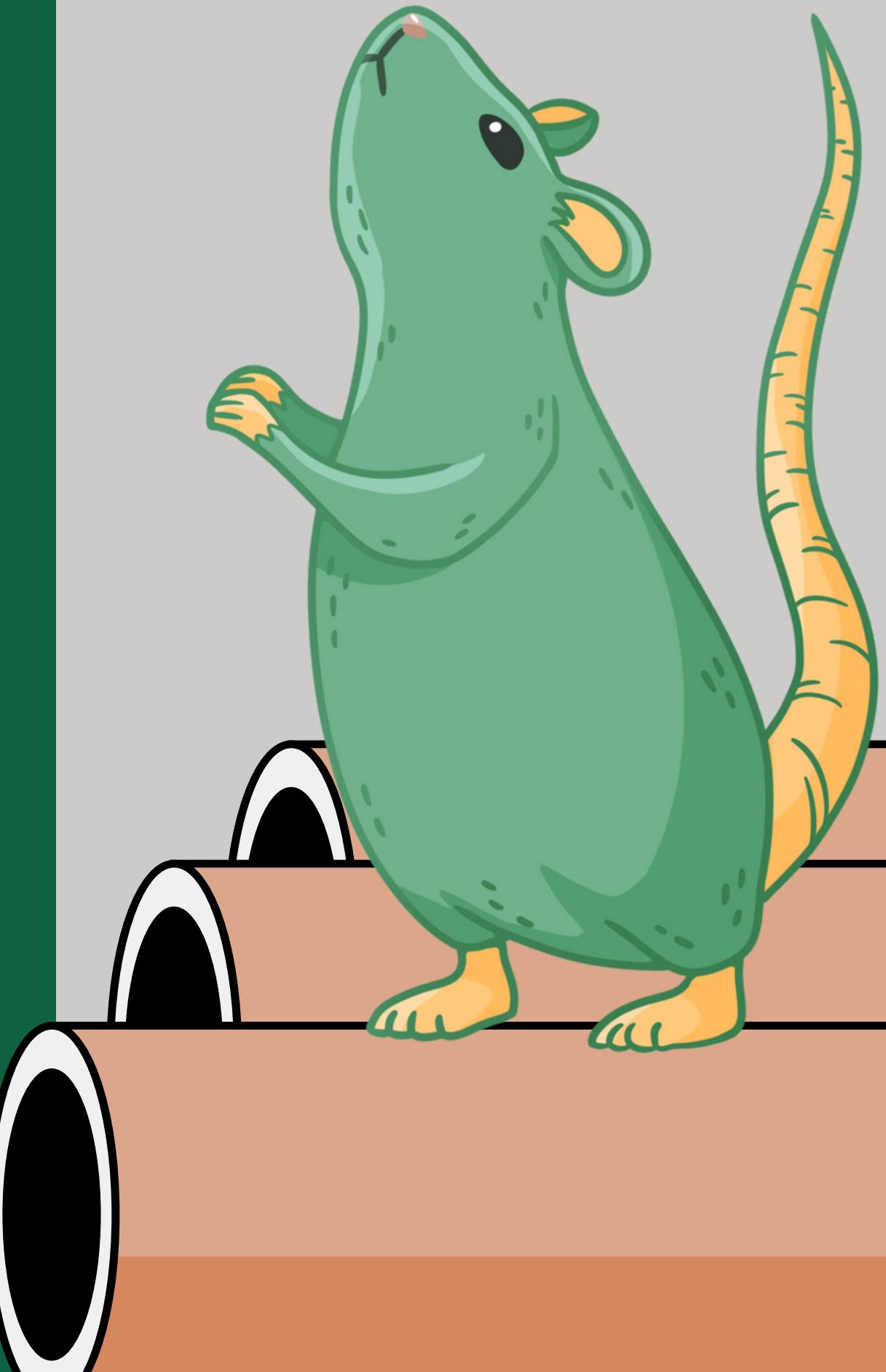
# MODEL PODSTAWOWY ZE ZMIENNYMI JAKOŚCIOWYMI SPOSTRZEŻENIA

**Problem:** Interesującą nas zmienną w "BodyWeight" była zmienna, której nie można zmierzyć ilościowo.

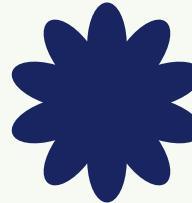
**Rozwiązanie:** Dodaliśmy zmienne do modelu tworząc zmienne fikcyjne (0-1) dla każdej kategorii.

**Problem:** Współliniowość.

**Rozwiązanie:** Wyłączenie jednej ze zmiennych, w naszym przypadku zmienna "b".



# MODEL PODSTAWOWY ZE ZMIENNYMI JAKOŚCIOWYMI



## UZASADNIENIE MERYTORYCZNE

- Zrozumienie wpływu różnych diet na przyrost i utratę masy ciała może pomóc w sformułowaniu zaleceń dietetycznych i interwencji dla poszczególnych osób i populacji.
- Testowanie na zwierzętach jest nieetyczne i jesteśmy temu przeciwni
- Uzyskane wyniki z badania na zwierzętach nie zawsze mogą być bezpośrednio przenoszone na ludzi.





# ANALIZA MODELU

- Wyjaśnienia modelu: 92% i 91,9% warjancji zmiennej zależnej!
- Istotność modelu:  $F = 998.2$  z wartością p-value mniejszą niż 0,001! NIESAMOWICIE!
- Mediana reszt wynosi 1.284
- Oszacowany współczynnik dla zmiennej a wynosi -220,989, co oznacza, że szczur na diecie a ważył średnio o 220,989 g mniej niż szczur nie będący na diecie podstawowej, utrzymując inne diety na stałym poziomie
- Współczynnik punktu przecięcia (484,705) to szacowana masa szczura, który był karmiony dietą podstawową.
- Oszacowany współczynnik dla zmiennej c wynosi 41,091, co oznacza, że szczur na diecie c ważył średnio 41,091 g więcej niż szczur nie będący na diecie c, utrzymując inne diety na stałym poziomie.

Call:

```
lm(formula = weight ~ a + c, data = BodyWeight)
```

Residuals:

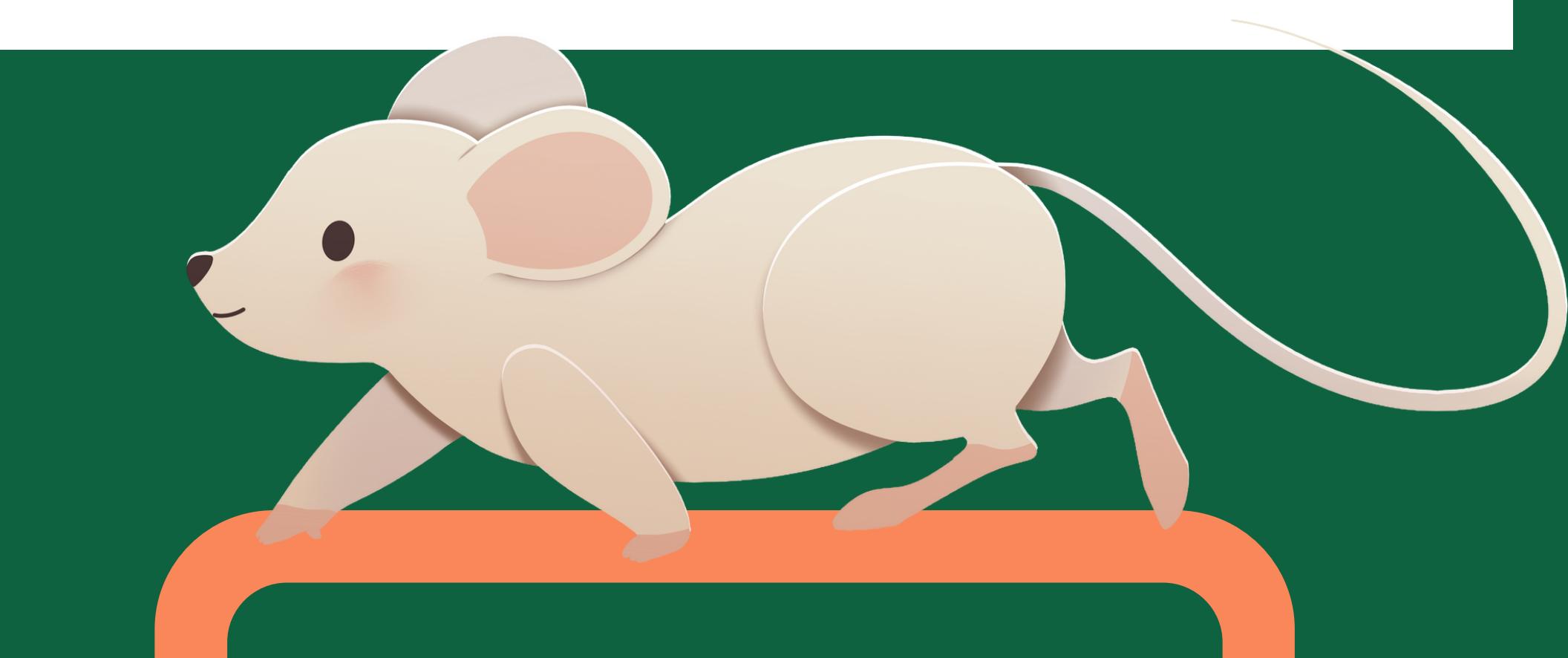
Min	1Q	Median	3Q	Max
-79.705	-19.026	1.284	11.287	143.295

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	484.705	5.445	89.022	< 2e-16 ***
a	-220.989	6.668	-33.139	< 2e-16 ***
c	41.091	7.700	5.336	2.94e-07 ***
---				

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.12 on 173 degrees of freedom  
Multiple R-squared: 0.9203, Adjusted R-squared: 0.9193  
F-statistic: 998.2 on 2 and 173 DF, p-value: < 2.2e-16



# MODEL PODSTAWOWY ZE ZMIENNYMI JAKOŚCIOWYMI

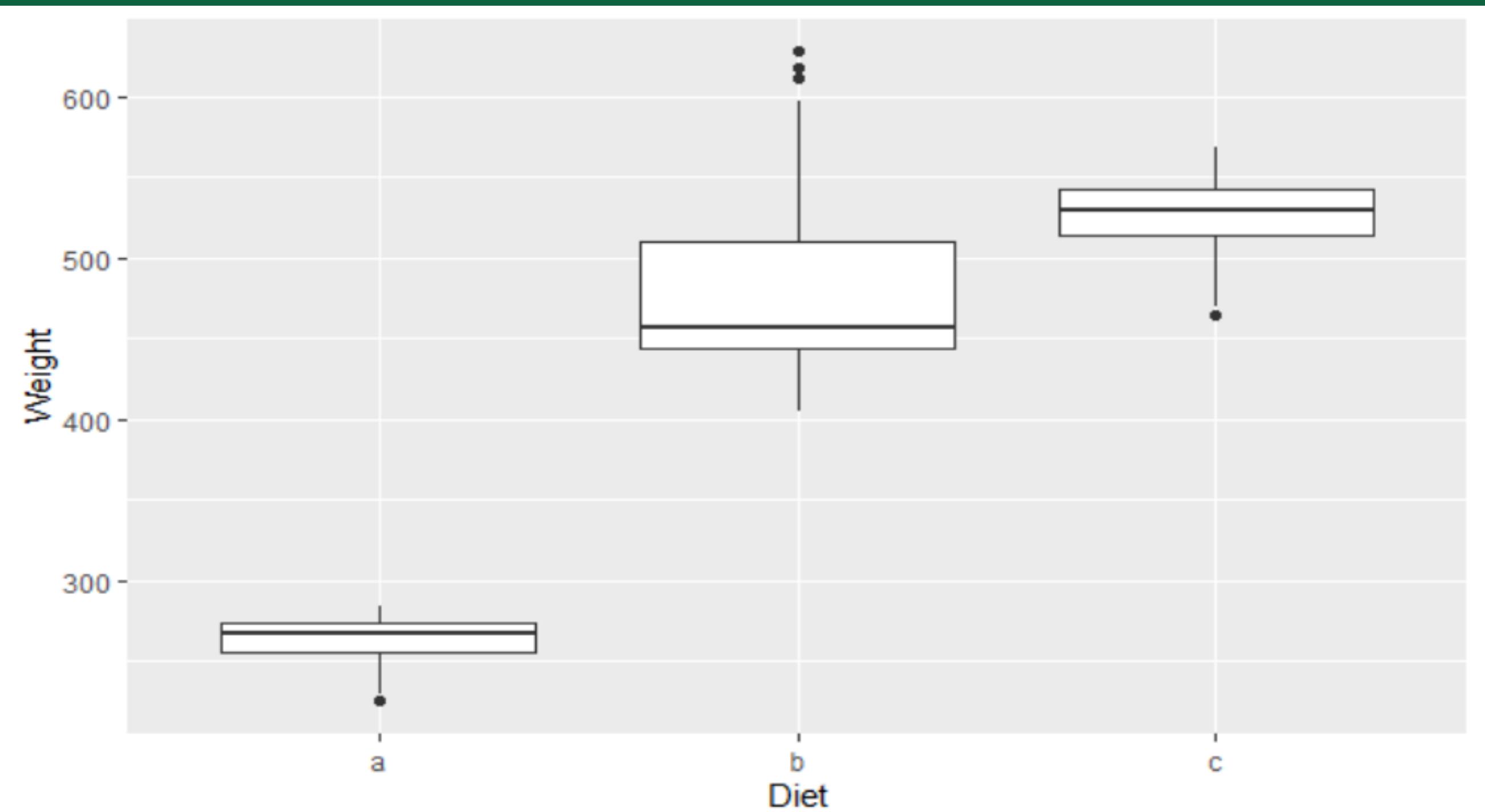
## PODSUMOWANIE

1 DIETA TYPU A / DIETA TYPU C

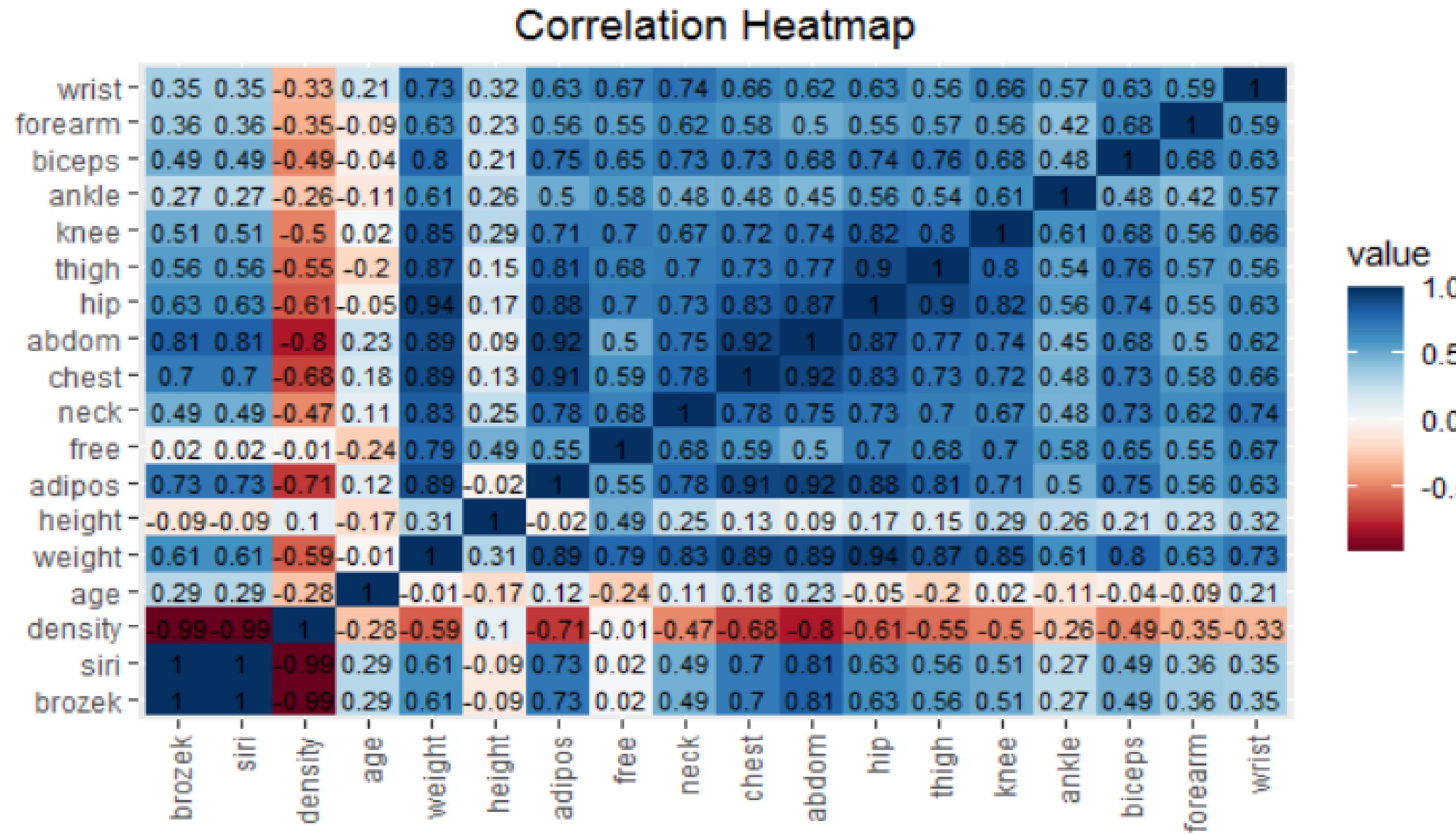
2 WARTOŚCI P DLA OBU ZMIENNYCH:  
\*\*\*

3  $R^2 = 0,9193$  /  $F = 988,2$  /  
P-VALUE < 0,001

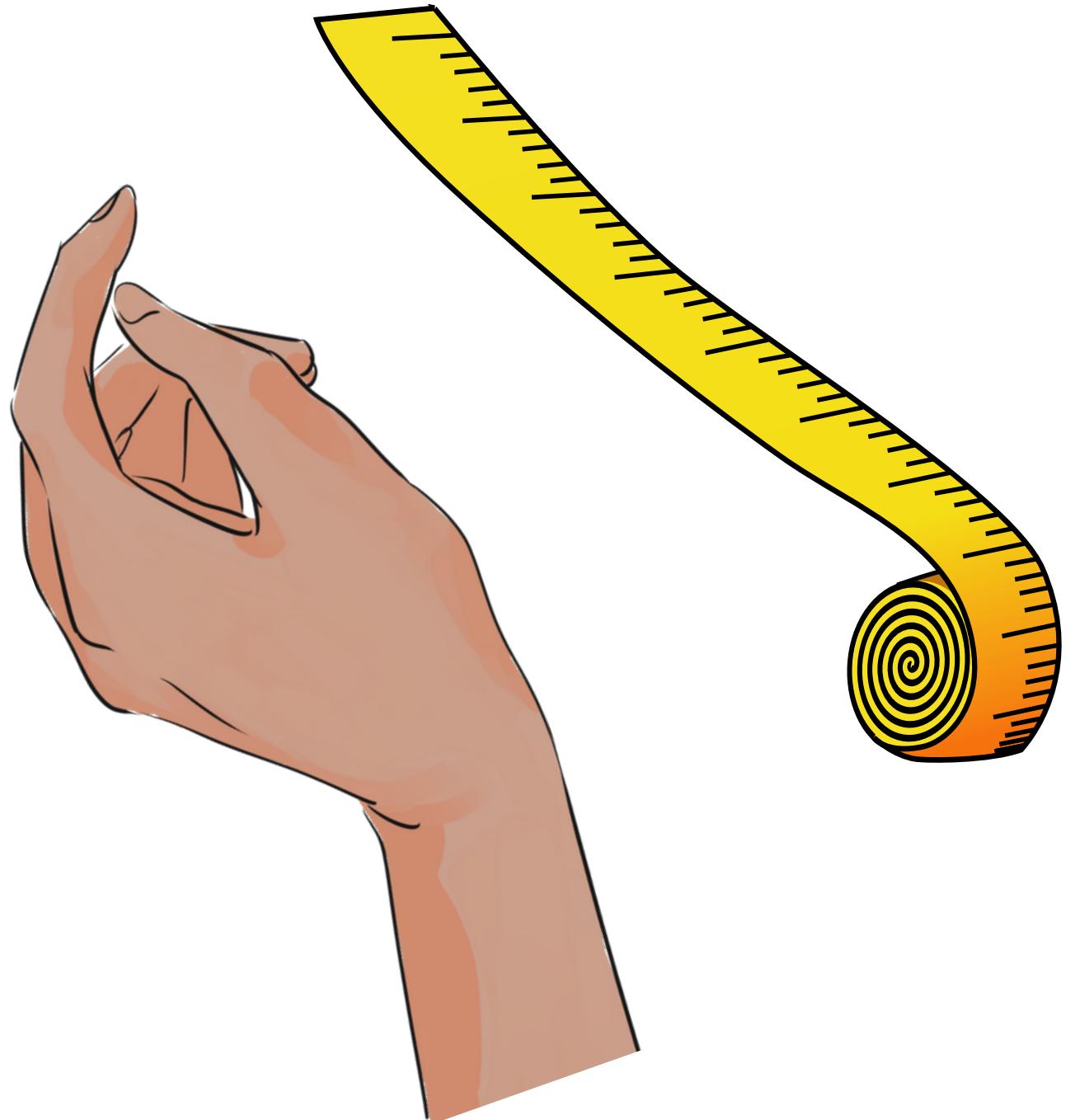
4 WYKRES



# Model podstawowy + współliniowość

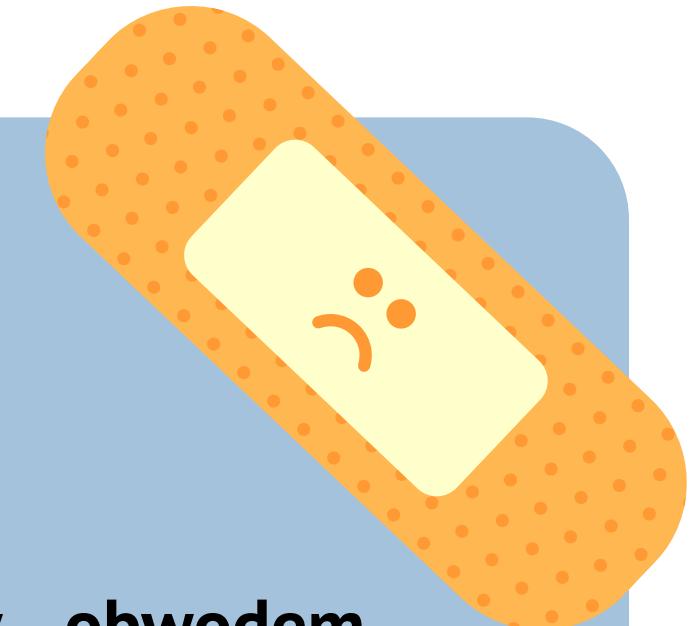


# Model podstawowy + współliniowość



## Uzasadnienie merytoryczne

Nasz model bada związek między obwodem nadgarstka a innymi pomiarami ciała, takimi jak wzrost, waga czy procent tkanki tłuszczowej. Obwód nadgarstka jest często używany jako wskaźnik wielkości szkieletu i może być używany do oszacowania rozkładu tkanki tłuszczowej i ryzyka różnych schorzeń.



# Model podstawowy + współliniowość

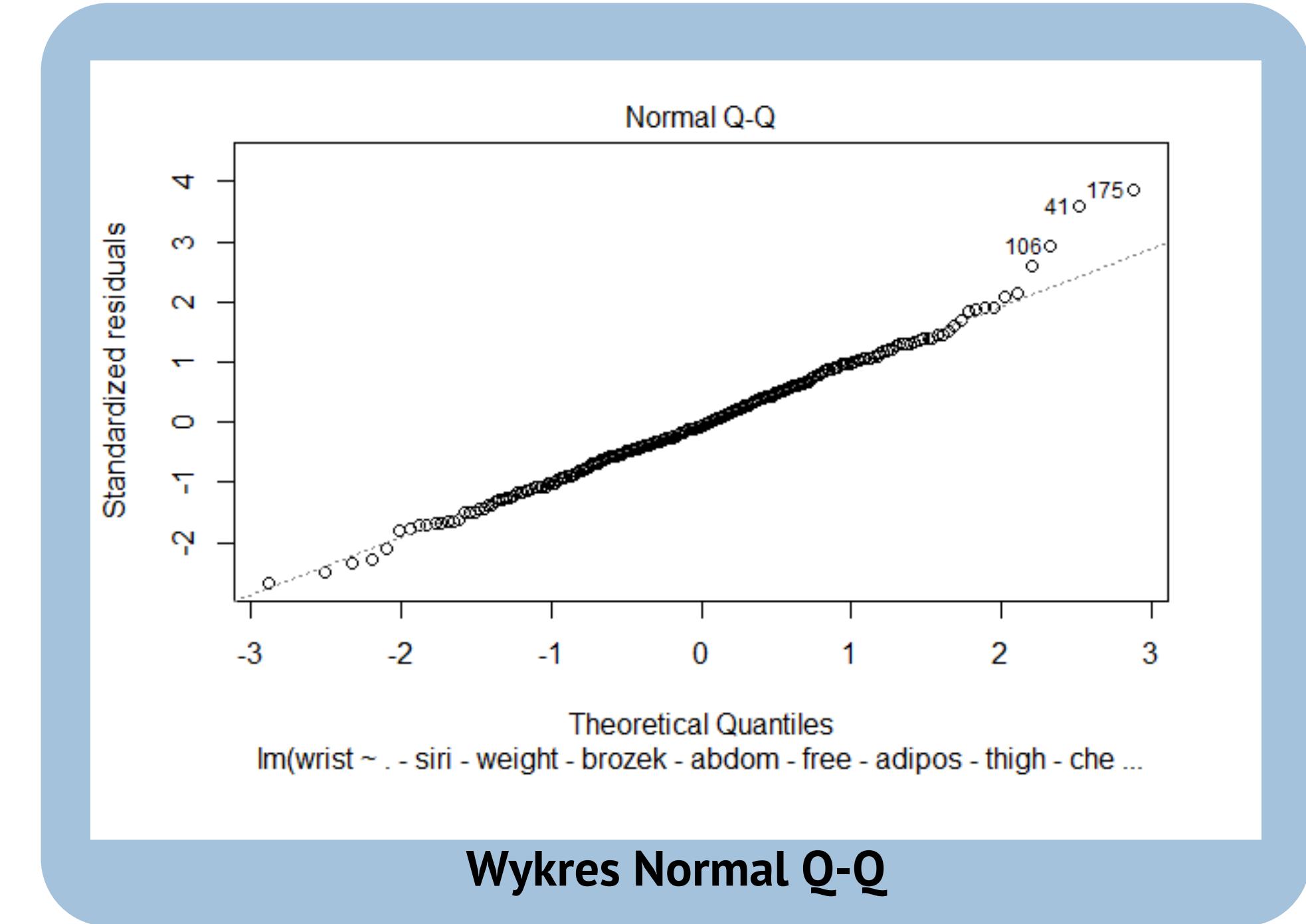
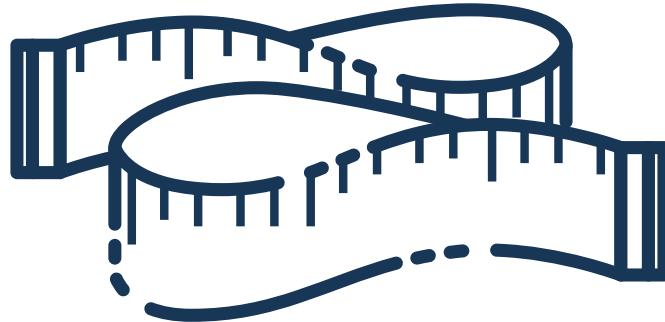
```
Call:  
lm(formula = wrist ~ . - siri - weight - brozek - abdom - free -  
    adipos - thigh - chest - hip - biceps, data = fat)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-1.19822 -0.32802 -0.03331  0.33014  1.79485  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) -6.296921  2.710024 -2.324 0.020971 *  
density       7.470883  2.199372  3.397 0.000796 ***  
age           0.019161  0.002820  6.796 8.21e-11 ***  
height        0.023284  0.009909  2.350 0.019587 *  
neck          0.159779  0.020688  7.723 2.93e-13 ***  
knee          0.076708  0.021848  3.511 0.000532 ***  
ankle         0.117659  0.024878  4.729 3.81e-06 ***  
forearm       0.083678  0.021699  3.856 0.000147 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.5155 on 244 degrees of freedom  
Multiple R-squared:  0.7036,   Adjusted R-squared:  0.6951  
F-statistic: 82.75 on 7 and 244 DF,  p-value: < 2.2e-16  
  
density      age      height      neck      knee      ankle      forearm  
1.654832  1.192556  1.244312  2.388866  2.622563  1.679276  1.815978
```

- wartości p wszystkich zmiennych niezależnych są statystycznie istotne (< 0,05)
- R-kwadrat wynoszący 0,7036 sugeruje, że model wyjaśnia około 70% wariancji zmiennej zależnej, natomiast bardziej skorygowany R-kwadrat mówi nam, że model wyjaśnia 69,51%
- statystyka F wynosi 82,75 z wartością p mniejszą niż 0,001, co wskazuje, że model jest wysoce istotny
- żadna z wartości VIF nie przekracza 5, a nawet maksymalna ma poniżej 3, co sugeruje, że w modelu nie ma poważnej współliniowości
- współczynniki pokazują szacunkową zmianę obwodu nadgarstka związaną ze wzrostem o jedną jednostkę w każdej zmiennej niezależnej

# Model podstawowy + współliniowość

## Podsumowanie

Oszacowania współczynników dla zmiennych w modelu pokazują, że prawie wszystkie mają statystycznie istotny związek z obwodem nadgarstka, podczas gdy wzrost jest marginalnie istotny. Model ma dobre dopasowanie (wielokrotny R-kwadrat: 0,7036), a skorygowany R-kwadrat sugeruje, że model może wyjaśnić 69,51% zmienności obwodu nadgarstka.



# Model z moderacją

## Opis

Zbior danych: "fat"

Zmienne wykorzystywanyymi w naszym modelu są:

- Z (moderatorem) jest zmienna "thigh",
- X - zmienna "chest"
- Y - zmienna "abdom".

## Uzasadnienie merytoryczne

Model sugeruje, że zarówno obwód klatki piersiowej, jak i obwód uda są pozytywnie związane z tłuszczem brzucha.

Przydatność: np. zdrowie i sprawność fizyczna, w których badacze lub praktycy są zainteresowani przewidywaniem tkanki tłuszczowej w jamie brzusznej na podstawie pomiarów fizycznych takich obszarów jak klatka piersiowa czy uda.

# Model z moderacją

```
Call:  
lm(formula = abdom ~ xc + zc + xc * zc, data = fat)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-11.7076 -2.6530  0.1556  2.7936  9.7125  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 92.322948   0.277933 332.177 < 2e-16 ***  
xc          0.968787   0.044285  21.876 < 2e-16 ***  
zc          0.405890   0.072019   5.636 4.72e-08 ***  
xc:zc       0.007242   0.003501   2.068   0.0396 *  
---  
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 4.033 on 248 degrees of freedom  
Multiple R-squared:  0.8618, Adjusted R-squared:  0.8601  
F-statistic: 515.3 on 3 and 248 DF, p-value: < 2.2e-16  
  
Call:  
lm(formula = abdom ~ xc + zc + xc * zc, data = fat)  
  
Coefficients:  
(Intercept)          xc          zc        xc:zc  
  92.322948    0.968787   0.405890   0.007242  
  
ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
Level of Significance = 0.05  
  
Call:  
gvlma(x = fitmod)
```

- Istotność.

Interpretacja:

- Jeśli obwód klatki piersiowej wzrośnie o jeden cm, model przewiduje, że obwód brzucha wzrośnie średnio o 0,968787 cm, zakładając, że wszystkie inne zmienne będą na stałym poziomie.
- Obwód brzucha wzrośnie średnio o 0,405890 cm, zakładając, że wszystkie inne zmienne będą na stałym poziomie.
- Współczynnik terminu interakcji, „xc:zc” = 0,007242. Przedstawia on zmianę zależności między obwodem klatki piersiowej, a obwodem brzucha, gdy uwzględniony zostanie obwód uda. (Znak dodatni wskazuje, że wpływ obwodu klatki piersiowej na obwód brzucha jest silniejszy, gdy obwód uda jest większy.)

# Model z moderacją

## Podsumowanie

Modelem regresji liniowej z:

- "abdom" jako zmienną zależną
- xc(scentrowany obwód klatki piersiowej),
- zc (scentrowany obwód uda),
- ich terminem interakcji **xc:zc** jako zmiennymi niezależnymi.

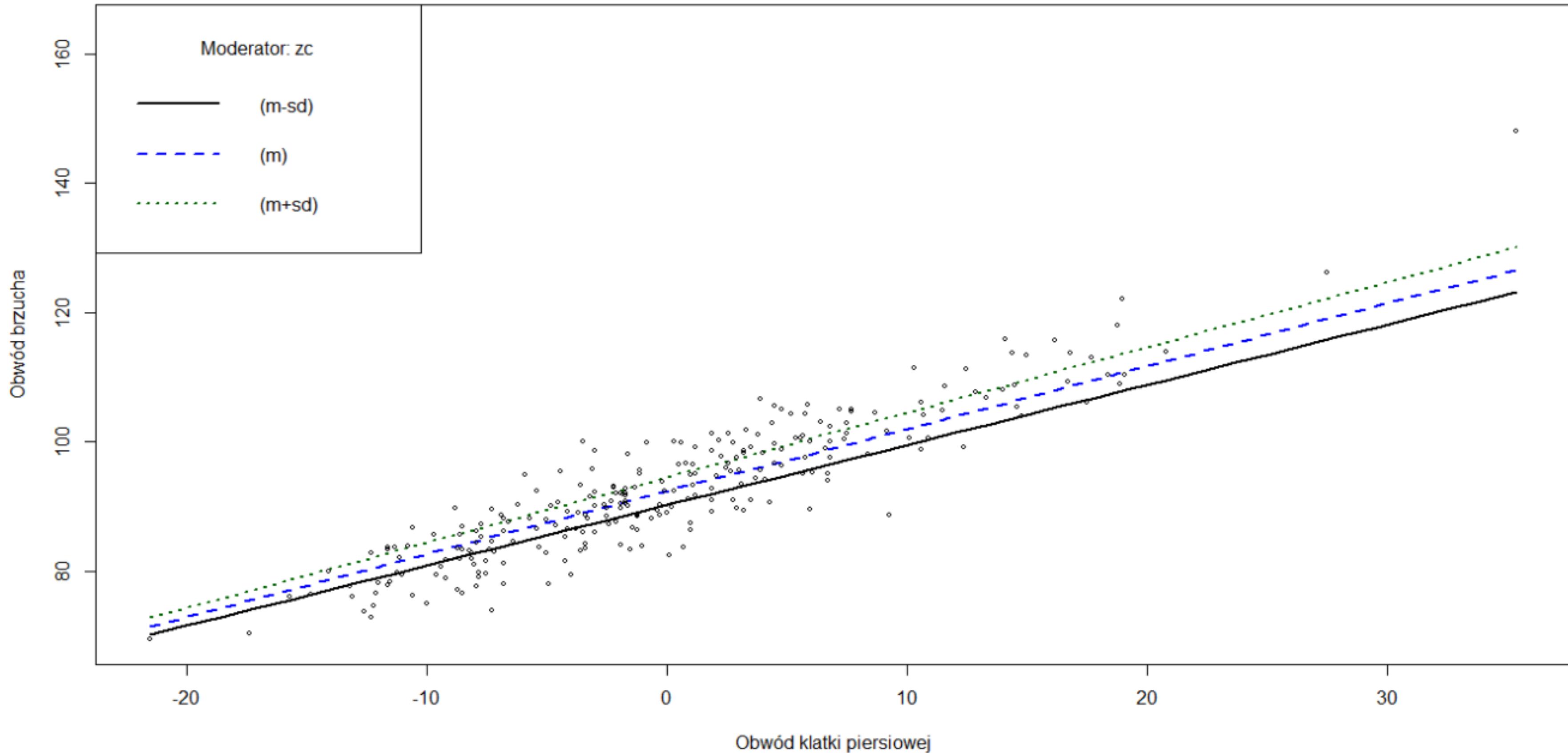
Celem modelu było przewidywanie obwodu brzucha na podstawie zmiennych xc i zc.

Akceptowalność modelu (86,18%)

	Value <dbl>	p-value <dbl>	Decision <chr>
Global Stat	3.651498723	0.4552194	Assumptions acceptable.
Skewness	0.385626529	0.5346073	Assumptions acceptable.
Kurtosis	0.001925178	0.9650026	Assumptions acceptable.
Link Function	1.768394047	0.1835810	Assumptions acceptable.
Heteroscedasticity	1.495552969	0.2213569	Assumptions acceptable.

5 rows

# Model z moderacją



**Dziękujemy za  
uwagę!**

