

# ЗВІТ

з лабораторної роботи №3  
з дисципліни «Аналіз даних»

Склад команди:  
Піковець Артем КМ-22

# 1 Вступ

Підвищений артеріальний тиск є одним із головних факторів ризику серцево-судинних захворювань. Це дослідження спрямоване на дослідження наступного питання:

- Як вісцеральний жир впливає на систолічний тиск людини?
- Гіпотеза: збільшення маси вісцерального жиру призводить до підвищення систолічного тиску.

## 2 Результати

### 2.1 Базова модель, моделі з контрольними змінними

Будемо досліджувати вплив збільшення вісцерального жиру на певний відсоток на систолічний тиск.

- Маємо натсупну базову модель:

$$y = \beta_0 + \beta_1 \ln(visceral\_fat)$$

Збільшення вісцерального жиру на  $\alpha\%$  пов'язано зі збільшенням систолічного тиску на  $\frac{\alpha}{100}\beta$  mmHg, де  $\beta$  - коефіцієнт при масі вісцерального жиру,  $\alpha$  - достатньо малий відсоток.

- Маса вісцерального жиру зростає з віком, при цьому вік має вплив на систолічний тиск. Тому варто контролювати вік:

$$y = \beta_0 + \beta_1 \ln(visceral\_fat) + \beta_2 age$$

- Маса вісцерального жиру відрізняється за статтю та расою, при цьому стать та раса можуть мати вплив на систолічний тиск. Тому варто їх контролювати<sup>1</sup>:

$$y = \beta_0 + \beta_1 \ln(visceral\_fat) + \beta_2 age + \beta_3 gender + \beta_4 race$$

- Маса вісцерального жиру корелює з розміром людини (нежирова маса, висота людини), при цьому розмір людини може мати вплив на систолічний тиск. Тому варто його контролювати<sup>2</sup>:

$$y = \beta_0 + \beta_1 \ln(visceral\_fat) + \beta_2 age + \beta_3 gender + \beta_4 race + \beta_5 \ln(lean\_mass) + \beta_6 height$$

- Маса вісцерального жиру корелює з масою інших видів жиру, маси яких також можуть мати вплив на систолічний тиск. Тому варто їх контролювати:

$$y = \beta_0 + \beta_1 \ln(visceral\_fat) + \beta_2 age + \beta_3 gender + \beta_4 race + \beta_5 \ln(lean\_mass) + \beta_6 height + \beta_7 \ln(subcutaneous\_fat) + \beta_8 \ln(gynoid\_fat)$$

---

<sup>1</sup>Для зручності не будемо відображати кодування категорійних змінних в окремі категорії

<sup>2</sup>Змінні які репрезентують маси будемо логаритмувати

Оцінимо вищенаведені моделі.

	<i>Dependent variable:</i>				
	systolic				
	(1)	(2)	(3)	(4)	(5)
log(visceral_fat_g)	9.950*** (0.148)	5.092*** (0.176)	4.943*** (0.181)	1.970*** (0.210)	2.018*** (0.250)
age		0.349*** (0.009)	0.361*** (0.009)	0.328*** (0.010)	0.320*** (0.010)
genderMale			3.752*** (0.195)	2.116*** (0.199)	1.907*** (0.306)
raceNon-Hispanic Asian			−0.318 (0.340)	0.371 (0.338)	0.401 (0.337)
raceNon-Hispanic Black			5.156*** (0.318)	3.514*** (0.317)	3.714*** (0.321)
raceNon-Hispanic White			−0.029 (0.274)	−0.356 (0.270)	−0.143 (0.274)
raceOther Hispanic			0.410 (0.366)	0.276 (0.357)	0.433 (0.359)
raceOther Race			1.665*** (0.443)	1.056** (0.425)	1.172*** (0.426)
log(lean_mass_g)				14.704*** (0.717)	15.215*** (0.793)
height_cm				−0.158*** (0.016)	−0.133*** (0.017)
log(subcutaneous_fat_g)					2.196*** (0.364)
log(gynoid_fat_g)					−3.633*** (0.586)
Constant	57.187*** (0.825)	74.871*** (0.858)	72.222*** (0.922)	−39.378*** (4.751)	−34.664*** (4.981)
Observations	16,788	16,788	16,788	16,788	16,788
Adjusted R <sup>2</sup>	0.221	0.299	0.335	0.362	0.363

---

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

У кожній з наведених моделей коефіцієнт при вісцеральному жиру статистично значущий та більше 0, але з додаванням контрольних змінних спостерігається його зменшення. Коефіцієнти при контрольних змінних також виявились статистично значущими. Значення  $R^2$  невелике.

## 2.2 Моделі з факторами взаємодії

- Вплив вісцерального жиру може бути різний при різних розмірах тіла, зокрема для людей з меншим зростом збільшення вісцерального жиру може більше впливати на тиск, ніж для людей з такою ж вагою але більшим зростом. Тому варто включити відповідні фактори взаємодії:

$$\begin{aligned} y = & \beta_0 + (\beta_1 + \beta_9 \ln(\text{lean\_mass}) + \beta_{10} \text{height}) \cdot \ln(\text{visceral\_fat}) \\ & + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{race} \\ & + \beta_5 \ln(\text{lean\_mass}) + \beta_6 \text{height} + \beta_7 \ln(\text{subcutaneous\_fat}) + \beta_8 \ln(\text{gynoid\_fat}) \end{aligned}$$

- Вплив вісцерального жиру може змінюватися з віком. Тому варто включити відповідний фактор взаємодії:

$$\begin{aligned} y = & \beta_0 + (\beta_1 + \beta_9 \ln(\text{lean\_mass}) + \beta_{10} \text{height} + \beta_{11} \text{age}) \cdot \ln(\text{visceral\_fat}) \\ & + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{race} \\ & + \beta_5 \ln(\text{lean\_mass}) + \beta_6 \text{height} + \beta_7 \ln(\text{subcutaneous\_fat}) + \beta_8 \ln(\text{gynoid\_fat}) \end{aligned}$$

- Вплив вісцерального жиру може бути різним для чоловіків та жінок. Тому варто включити відповідний фактор взаємодії, при цьому додавши фактори взаємодії статі з контрольними змінними що корелюють з вісцеральним жиром та також можуть мати різний вплив на тиск в залежності від статі:

$$\begin{aligned} y = & \beta_0 + (\beta_1 + \beta_9 \ln(\text{lean\_mass}) + \beta_{10} \text{height} + \beta_{11} \text{age} + \beta_{12} \text{gender}) \cdot \ln(\text{visceral\_fat}) \\ & + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{race} \\ & + \beta_5 \ln(\text{lean\_mass}) + \beta_6 \text{height} + \beta_7 \ln(\text{subcutaneous\_fat}) + \beta_8 \ln(\text{gynoid\_fat}) \\ & + (\beta_{14} \text{age} + \beta_{15} \ln(\text{lean\_mass}) + \beta_{16} \text{height} + \beta_{17} \ln(\text{subcutaneous\_fat}) \\ & + \beta_{18} \ln(\text{gynoid\_fat})) \cdot \text{gender} \end{aligned}$$

- Вплив віку на тиск відрізняється за расою, особливо між темношкірими та іншими. Додамо відповідний фактор взаємодії:

$$\begin{aligned} y = & \beta_0 + (\beta_1 + \beta_9 \ln(\text{lean\_mass}) + \beta_{10} \text{height} + \beta_{11} \text{age} + \beta_{12} \text{gender}) \cdot \ln(\text{visceral\_fat}) \\ & + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{race} \\ & + \beta_5 \ln(\text{lean\_mass}) + \beta_6 \text{height} + \beta_7 \ln(\text{subcutaneous\_fat}) + \beta_8 \ln(\text{gynoid\_fat}) \\ & + (\beta_{14} \text{age} + \beta_{15} \ln(\text{lean\_mass}) + \beta_{16} \text{height} + \beta_{17} \ln(\text{subcutaneous\_fat}) \\ & + \beta_{18} \ln(\text{gynoid\_fat})) \cdot \text{gender} \\ & + \beta_{19} \text{age} \cdot \text{is\_black} \end{aligned}$$

Оцінимо вищенаведені моделі.

	<i>Dependent variable:</i>			
	systolic			
	(1)	(2)	(3)	(4)
log(visceral_fat_g)	2.018*** (0.250)	-16.553*** (5.105)	-7.956 (5.889)	4.917 (6.184)
age	0.320*** (0.010)	0.321*** (0.010)	0.081 (0.093)	-0.056 (0.094)
genderMale	1.907*** (0.306)	2.298*** (0.316)	2.409*** (0.318)	-44.190*** (10.404)
raceNon-Hispanic Asian	0.401 (0.337)	0.485 (0.338)	0.525 (0.338)	0.793** (0.337)
raceNon-Hispanic Black	3.714*** (0.321)	3.882*** (0.322)	3.821*** (0.323)	-0.618 (0.472)
raceNon-Hispanic White	-0.143 (0.274)	-0.027 (0.276)	-0.034 (0.276)	0.243 (0.275)
raceOther Hispanic	0.433 (0.359)	0.504 (0.359)	0.486 (0.359)	0.637* (0.357)
raceOther Race	1.172*** (0.426)	1.285*** (0.427)	1.271*** (0.427)	1.314*** (0.425)
log(lean_mass_g)	15.215*** (0.793)	-4.395 (4.444)	1.553 (4.844)	10.835*** (5.127)
height_cm	-0.133*** (0.017)	0.452*** (0.124)	0.415*** (0.124)	0.260** (0.127)
log(subcutaneous_fat_g)	2.196*** (0.364)	2.504*** (0.389)	2.526*** (0.389)	1.589* (0.902)
log(gynoid_fat_g)	-3.633*** (0.586)	-3.819*** (0.616)	-3.597*** (0.625)	-4.612*** (0.964)
log(visceral_fat_g):log(lean_mass_g)		3.271*** (0.761)	2.243*** (0.837)	0.699 (0.865)
log(visceral_fat_g):height_cm		-0.100*** (0.022)	-0.094*** (0.022)	-0.075*** (0.022)

log(visceral_fat_g):age			0.041*** (0.015)	0.068*** (0.016)
log(visceral_fat_g):genderMale				1.654** (0.716)
age:genderMale				-0.179*** (0.021)
genderMale:log(lean_mass_g)				2.183 (1.612)
genderMale:height_cm				0.053 (0.036)
genderMale:log(subcutaneous_fat_g)				0.416 (1.115)
genderMale:log(gynoid_fat_g)				0.787 (1.422)
age:is_black				0.167*** (0.018)
Constant	-34.664*** (4.981)	77.937*** (29.396)	26.137 (33.987)	-29.933 (36.039)
Observations	16,788	16,788	16,788	16,788
Adjusted R <sup>2</sup>	0.363	0.364	0.364	0.374

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Тепер складніше сказати чи коефіцієнт при вісцеральному жиру більше 0, бо він залежить від значень інших змінних, дане питання буде розглянуто у фінальній моделі. Значення  $R^2$  трохи підвищилось.

## 2.3 Моделі з поліномами

Розглядалися наступні моделі з поліномами від змінної віку:

$$\begin{aligned}
y = & \dots + (\beta_1 age + \beta_2 age^2 + \dots) \\
& + (\beta_{v1} age + \beta_{v2} age^2 + \dots) \cdot \ln(visceral\_fat) \\
& + (\beta_{g1} age + \beta_{g2} age^2 + \dots) \cdot gender \\
& + (\beta_{r1} age + \beta_{r2} age^2 + \dots) \cdot race + \dots
\end{aligned}$$

Було обрано модель де коефіцієнти при ступенях віку були статистично значущі та не спостерігалось значного підвищення  $R^2$  при додаванні вищих ступенів.

	<i>Dependent variable:</i>		
	systolic		
	(1)	(2)	(3)
log(visceral_fat_g)	4.917 (6.184)	0.464 (6.488)	-1.193 (6.512)
age	-0.056 (0.094)	-1.134*** (0.408)	2.634** (1.277)
I(age^2)		0.007 (0.008)	-0.135*** (0.046)
I(age^3)		0.0003*** (0.0001)	0.002*** (0.001)
I(age^4)			-0.00001 (0.00001)
genderMale	-44.190*** (10.404)	-20.486* (10.764)	-17.668 (10.799)
raceNon-Hispanic Asian	0.793** (0.337)	0.966*** (0.334)	0.950*** (0.334)
raceNon-Hispanic Black	-0.618 (0.472)	-3.089*** (0.933)	-1.395 (3.859)
raceNon-Hispanic White	0.243 (0.275)	0.310 (0.274)	0.337 (0.275)
raceOther Hispanic	0.637* (0.357)	0.660* (0.354)	0.656* (0.354)
raceOther Race	1.314*** (0.425)	1.356*** (0.424)	1.387*** (0.425)
log(lean_mass_g)	10.835** (5.127)	10.656** (5.229)	8.728* (5.291)
height_cm	0.260** (0.127)	0.298** (0.129)	0.187 (0.134)
log(subcutaneous_fat_g)	1.589* (0.902)	0.831 (0.900)	1.051 (0.904)
log(gynoid_fat_g)	-4.612*** (0.964)	-3.983*** (0.967)	-4.226*** (0.971)

log(visceral_fat_g):log(lean_mass_g)	0.699 (0.865)	1.043 (0.890)	1.342 (0.901)
log(visceral_fat_g):height_cm	-0.075*** (0.022)	-0.080*** (0.023)	-0.062*** (0.024)
log(visceral_fat_g):age	0.068*** (0.016)	0.279*** (0.075)	-0.239 (0.243)
log(visceral_fat_g):I(age^2)		-0.004*** (0.001)	0.013 (0.008)
log(visceral_fat_g):I(age^3)			-0.0002** (0.0001)
log(visceral_fat_g):genderMale	1.654** (0.716)	2.051*** (0.721)	1.740** (0.729)
age:genderMale	-0.179*** (0.021)	0.928*** (0.283)	0.088 (0.724)
I(age^2):genderMale		-0.028*** (0.009)	0.021 (0.039)
I(age^3):genderMale		0.0002** (0.0001)	-0.001 (0.001)
I(age^4):genderMale			0.00001 (0.00001)
genderMale:log(lean_mass_g)	2.183 (1.612)	-0.773 (1.655)	-0.606 (1.674)
genderMale:height_cm	0.053 (0.036)	0.015 (0.037)	0.023 (0.037)
genderMale:log(subcutaneous_fat_g)	0.416 (1.115)	1.063 (1.113)	0.890 (1.114)
genderMale:log(gynoid_fat_g)	0.787 (1.422)	0.227 (1.423)	0.461 (1.426)
age:is_black	0.167*** (0.018)	0.391*** (0.080)	0.304 (0.698)
I(age^2):is_black		-0.004*** (0.001)	-0.011 (0.041)



I(age^3):is_black			0.0004 (0.001)
I(age^4):is_black			-0.00000 (0.00001)
Constant	-29.933 (36.039)	-23.137 (37.451)	-14.096 (37.657)

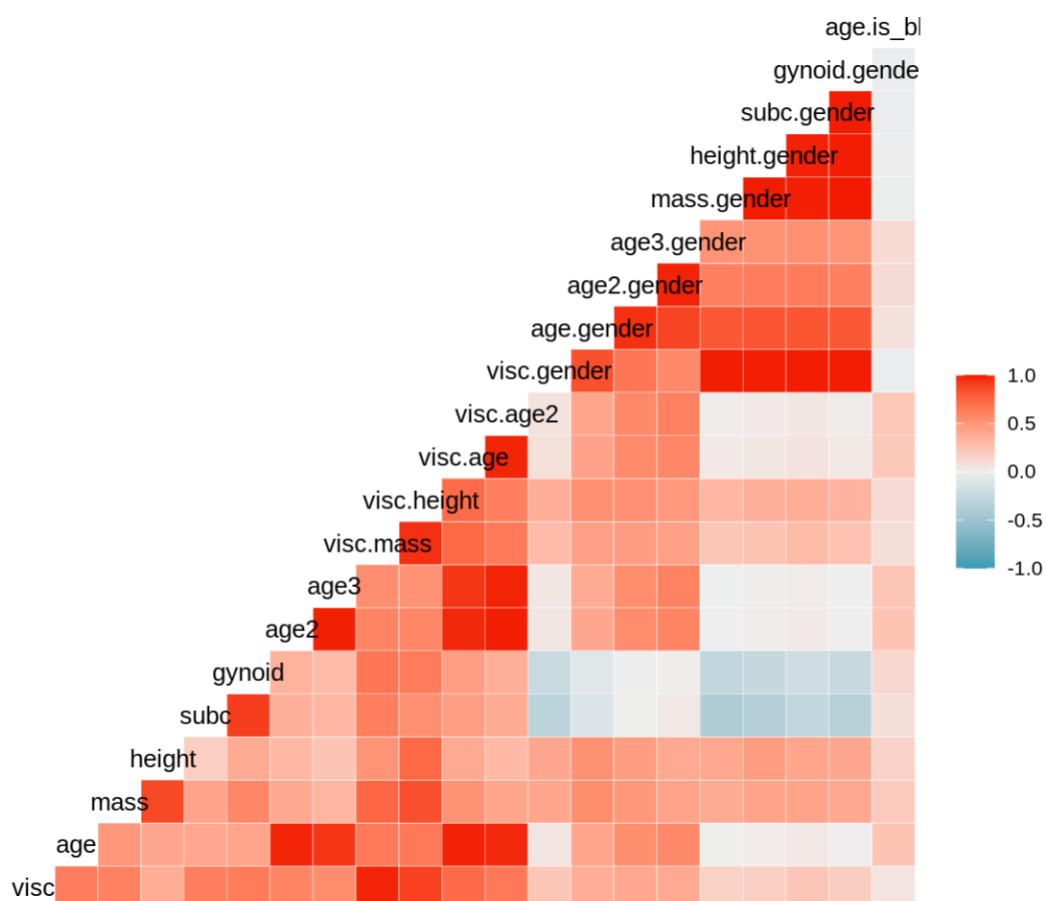
Observations	16,788	16,788	16,788
Adjusted R <sup>2</sup>	0.374	0.381	0.382

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

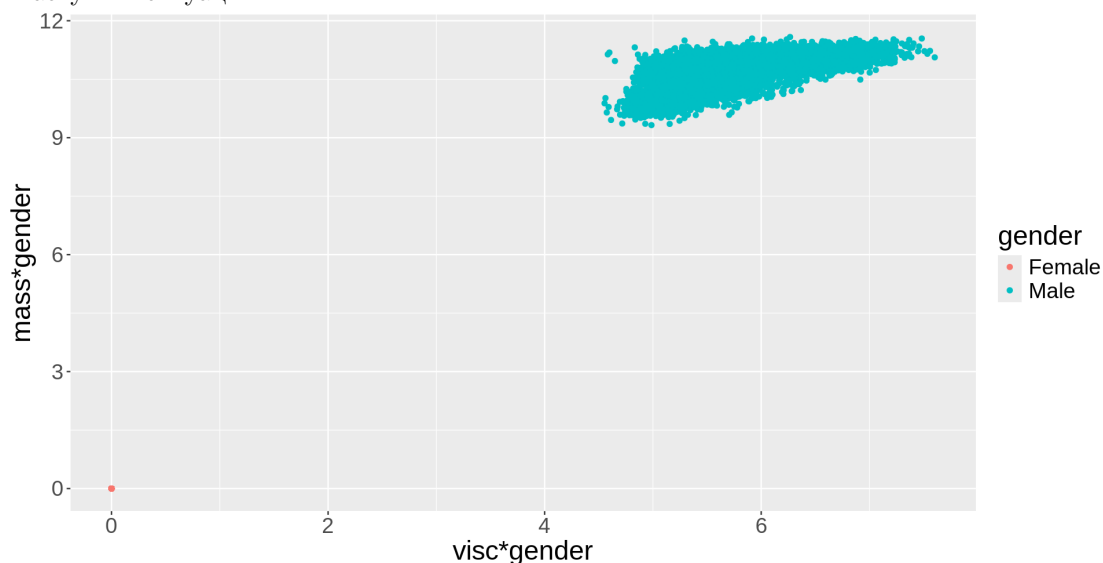
Обрана модель - модель (2).

## 2.4 Фінальна модель

Кореляції між регресорами



Спостерігаються кореляції близькі до 1. Але такі кореляції спостерігаються тільки між ступенями однієї змінної, або між змінною та фактором взаємодії з цією змінною, або ж при наступній ситуації:



Наврядчи це призведе до проблем обчислень OLS-оцінки, тому було вирішено відповідні змінні не прибрати.

### Коефіцієнт при вісцеральному жиру та його інтерпретація

Маємо наступну оцінку фінальної моделі:

$$y = (0.464 + 1.043 \cdot \ln(\text{lean\_mass}) - 0.080 \cdot \text{height} + 0.279 \cdot \text{age} - 0.0041 \cdot \text{age}^2 + 2.051 \cdot \text{genderMale}) \cdot \ln(\text{visceral\_fat}) + \dots$$

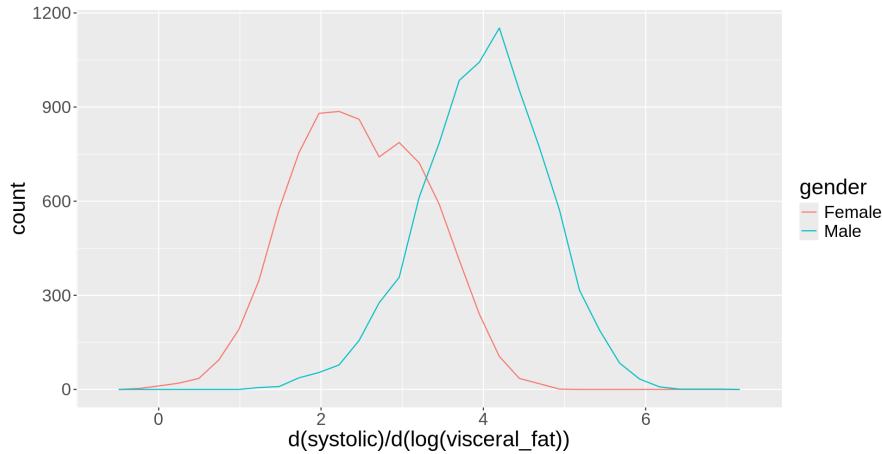
log(visceral_fat_g)	0.464 (6.488)
log(visceral_fat_g):log(lean_mass_g)	1.043 (0.890)
log(visceral_fat_g):height_cm	-0.080*** (0.023)
log(visceral_fat_g):age	0.279*** (0.075)
log(visceral_fat_g):I(age^2)	-0.004*** (0.001)
log(visceral_fat_g):genderMale	2.051***

(0.721)

Note:

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

Наведемо розподіл коефіцієнта при логаритму маси вісцерального жиру для даної вибірки людей.



Майже для всіх людей з вибірки даний коефіцієнт є позитивним. Тобто при збільшенні маси вісцерального жиру підвищується систолічний тиск.

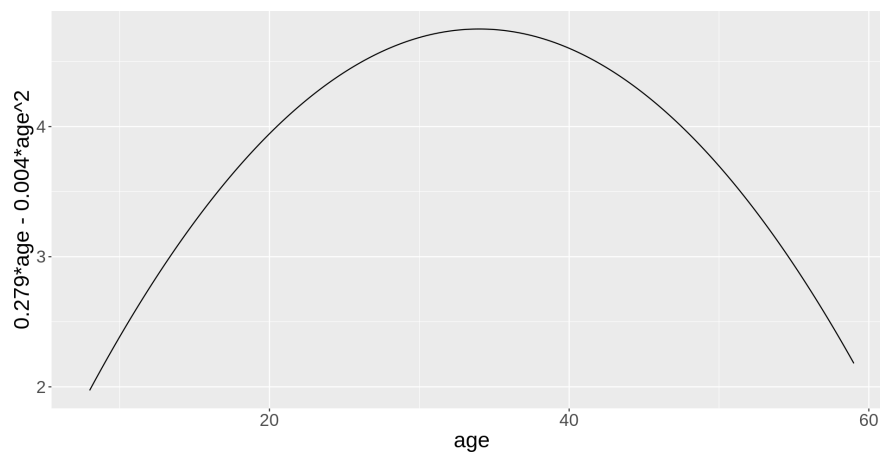
	$\text{median}(\beta_{\ln(\text{visceral\_fat})})$
Female	2.45
Male	4.04

Але підвищення систолічного тиску досить мале:

- Для жінок збільшення маси вісцерального жиру на 10% підвищує систолічний тиск на 0.245 mmHg.
- Для чоловіків збільшення маси вісцерального жиру на 10% підвищує систолічний тиск на 0.404 mmHg.

Залежність коефіцієнту при вісцеральному жиру від інших змінних:

- Коефіцієнт при нежировій масі не є статистично значущим. Мабуть вплив вісцерального жиру на систолічний тиск не залежить від маси людини.
- Коефіцієнт при зросту людини є статистично значущим та від'ємним. Тобто при збільшенні вісцерального жиру підвищення систолічного тиску є більшим для людей меншого зросту.
- Протестувавши гіпотезу  $H_0 : \beta_{age} = \beta_{age^2} = 0$  отримаємо мале р-значення:  $p = 0.00074$ . Тобто вплив вісцерального жиру на систолічний тиск залежить від віку людини, причому нелінійним чином:



- Коефіцієнт при статі є статистично значущим. Контролюючи нежирову масу, зріст та вік, для чоловіка коефіцієнт при вісцеральному жиру буде на 2.051 більше ніж для жінки.

### 3 Висновки

- Згідно побудованій регресійній моделі виявлено що збільшення маси вісцерального жиру призводить до незначного підвищення систолічного тиску.
- Вплив маси вісцерального жиру залежав від інших змінних, таких як вік та стать людини.
- Значення  $R^2_{adj.} = 0.381$  виявилось малим, можуть бути невраховані важливі змінні які корелюють з масою вісцерального жиру. Серед них можуть бути дієта людини, куріння, вживання алкоголю, фізична активність, рівень стресу.

### 4 Використані джерела

[1] Лекції 6, 7, 8, 9 з дисципліни "Аналіз даних"