

# Лабораторна робота 3

## Регресійний аналіз

Піковець Артем КМ-22

# Дослідницьке питання

- Питання: Як кількість вісцерального жиру впливає на систолічний тиск людини?
- Гіпотеза: Збільшення маси вісцерального жиру призводить до підвищення систолічного тиску.

- Дані взяті з NHANES за роки 2011-2018.
- Були взяті базові демографічні змінні, та зміні що стосуються артеріального тиску та розподілу жирової й нежирової маси в тілі людини.
- Отримана вибірка містить 16788 людей.

# Базова модель, моделі з контрольними змінними

- Базова модель:

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

# Базова модель, моделі з контрольними змінними

- Базова модель:

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

- Вік: Маса вісцерального жиру корелює з віком.

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

# Базова модель, моделі з контрольними змінними

- Базова модель:

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

- Вік: Маса вісцерального жиру корелює з віком.

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

- Стать та раса: маса вісцерального жиру відрізняється за статтю та расою.

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

# Базова модель, моделі з контрольними змінними

- Маса вісцерального жиру корелює з розміром людини (нежирова маса, висота людини).

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

# Базова модель, моделі з контрольними змінними

- Маса вісцерального жиру корелює з розміром людини (нежирова маса, висота людини).

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

- Маса вісцерального жиру корелює з масою інших видів жиру.

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



# Оцінка моделей I

|                        | <i>Dependent variable:</i> |                     |                     |                     |                     |
|------------------------|----------------------------|---------------------|---------------------|---------------------|---------------------|
|                        | systolic                   |                     |                     |                     |                     |
|                        | (1)                        | (2)                 | (3)                 | (4)                 | (5)                 |
| log(visceral_fat_g)    | 9.950***<br>(0.148)        | 5.092***<br>(0.176) | 4.943***<br>(0.181) | 1.970***<br>(0.210) | 2.018***<br>(0.250) |
| age                    |                            | 0.349***<br>(0.009) | 0.361***<br>(0.009) | 0.328***<br>(0.010) | 0.320***<br>(0.010) |
| genderMale             |                            |                     | 3.752***<br>(0.195) | 2.116***<br>(0.199) | 1.907***<br>(0.306) |
| raceNon-Hispanic Asian |                            |                     | −0.318<br>(0.340)   | 0.371<br>(0.338)    | 0.401<br>(0.337)    |
| raceNon-Hispanic Black |                            |                     | 5.156***<br>(0.318) | 3.514***<br>(0.317) | 3.714***<br>(0.321) |
| raceNon-Hispanic White |                            |                     | −0.029<br>(0.274)   | −0.356<br>(0.270)   | −0.143<br>(0.274)   |
| raceOther Hispanic     |                            |                     | 0.410<br>(0.366)    | 0.276<br>(0.357)    | 0.433<br>(0.359)    |

# Оцінка моделей II

|                         |                      |                      |                      |                       |                       |
|-------------------------|----------------------|----------------------|----------------------|-----------------------|-----------------------|
| raceOther Race          |                      |                      | 1.665***<br>(0.443)  | 1.056**<br>(0.425)    | 1.172***<br>(0.426)   |
| log(lean_mass_g)        |                      |                      |                      | 14.704***<br>(0.717)  | 15.215***<br>(0.793)  |
| height_cm               |                      |                      |                      | -0.158***<br>(0.016)  | -0.133***<br>(0.017)  |
| log(subcutaneous_fat_g) |                      |                      |                      |                       | 2.196***<br>(0.364)   |
| log(gynoid_fat_g)       |                      |                      |                      |                       | -3.633***<br>(0.586)  |
| Constant                | 57.187***<br>(0.825) | 74.871***<br>(0.858) | 72.222***<br>(0.922) | -39.378***<br>(4.751) | -34.664***<br>(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

# Моделі з факторами взаємодії

- В попередніх моделях припускалось що маса вісцерального жиру не корелює з іншими неврахованими змінними які можуть мати вплив на тиск.

# Моделі з факторами взаємодії

- В попередніх моделях припускалось що маса вісцерального жиру не корелює з іншими неврахованими змінними які можуть мати вплив на тиск.
- Припускаємо що отримані фактори взаємодії не корелюють з іншими неврахованими змінними які можуть мати вплив на тиск.

# Моделі з факторами взаємодії

- В попередніх моделях припускалось що маса вісцерального жиру не корелює з іншими неврахованими змінними які можуть мати вплив на тиск.
- Припускаємо що отримані фактори взаємодії не корелюють з іншими неврахованими змінними які можуть мати вплив на тиск.
- Вплив вісцерального жиру може бути різний при різних розмірах тіла.

$$y = \beta_0 + (\beta_1 + \beta_9 \ln(\text{lean\_mass}) + \beta_{10} \text{height}) \cdot \ln(\text{visceral\_fat}) + \dots$$

# Моделі з факторами взаємодії

- В попередніх моделях припускалось що маса вісцерального жиру не корелює з іншими неврахованими змінними які можуть мати вплив на тиск.
- Припускаємо що отримані фактори взаємодії не корелюють з іншими неврахованими змінними які можуть мати вплив на тиск.
- Вплив вісцерального жиру може бути різний при різних розмірах тіла.

$$y = \beta_0 + (\beta_1 + \beta_9 \ln(\text{lean\_mass}) + \beta_{10} \text{height}) \cdot \ln(\text{visceral\_fat}) + \dots$$

- Вплив вісцерального жиру може змінюватися з віком.

$$y = (\dots + \beta_{11} \text{age}) \cdot \ln(\text{visceral\_fat}) + \dots$$

# Моделі з факторами взаємодії

- Вплив вісцерального жиру може бути різним для чоловіків та жінок.

$$\begin{aligned} y = & (\dots + \beta_{12} \text{gender}) \cdot \ln(\text{visceral\_fat}) \\ & + (\beta_{14} \text{age} + \beta_{15} \ln(\text{lean\_mass}) + \beta_{16} \text{height} \\ & + \beta_{17} \ln(\text{subcutaneous\_fat}) + \beta_{18} \ln(\text{gynoid\_fat})) \cdot \text{gender} + \dots \end{aligned}$$

# Моделі з факторами взаємодії

- Вплив вісцерального жиру може бути різним для чоловіків та жінок.

$$\begin{aligned} y = & (\dots + \beta_{12}gender) \cdot \ln(visceral\_fat) \\ & + (\beta_{14}age + \beta_{15} \ln(lean\_mass) + \beta_{16}height \\ & + \beta_{17} \ln(subcutaneous\_fat) + \beta_{18} \ln(gynoid\_fat)) \cdot gender + \dots \end{aligned}$$

- Вплив віку на тиск відрізняється за расою, особливо між темношкірими та іншими.

$$y = \dots + \beta_{19}age \cdot is\_black$$



# Оцінка моделей I

|                                      | <i>Dependent variable:</i> |                       |                      |                      |
|--------------------------------------|----------------------------|-----------------------|----------------------|----------------------|
|                                      | systolic                   |                       |                      |                      |
|                                      | (1)                        | (2)                   | (3)                  | (4)                  |
| log(visceral_fat_g)                  | 2.018***<br>(0.250)        | -16.553***<br>(5.105) | -7.956<br>(5.889)    | 4.917<br>(6.184)     |
| log(visceral_fat_g):log(lean_mass_g) |                            | 3.271***<br>(0.761)   | 2.243***<br>(0.837)  | 0.699<br>(0.865)     |
| log(visceral_fat_g):height_cm        |                            | -0.100***<br>(0.022)  | -0.094***<br>(0.022) | -0.075***<br>(0.022) |
| log(visceral_fat_g):age              |                            |                       | 0.041***<br>(0.015)  | 0.068***<br>(0.016)  |
| log(visceral_fat_g):genderMale       |                            |                       |                      | 1.654**<br>(0.716)   |
| age:genderMale                       |                            |                       |                      | -0.179***<br>(0.021) |
| genderMale:log(lean_mass_g)          |                            |                       |                      | 2.183<br>(1.612)     |

# Оцінка моделей II

|                                    |                     |
|------------------------------------|---------------------|
| genderMale:height_cm               | 0.053<br>(0.036)    |
| genderMale:log(subcutaneous_fat_g) | 0.416<br>(1.115)    |
| genderMale:log(gynoid_fat_g)       | 0.787<br>(1.422)    |
| age:is_black                       | 0.167***<br>(0.018) |

|                         |        |        |        |        |
|-------------------------|--------|--------|--------|--------|
| Observations            | 16,788 | 16,788 | 16,788 | 16,788 |
| Adjusted R <sup>2</sup> | 0.363  | 0.364  | 0.364  | 0.374  |

- Розглядались 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.

# Оцінка моделей I

|                              | <i>Dependent variable:</i> |                       |                       |
|------------------------------|----------------------------|-----------------------|-----------------------|
|                              | systolic                   |                       |                       |
|                              | (1)                        | (2)                   | (3)                   |
| age                          | −0.056<br>(0.094)          | −1.134***<br>(0.408)  | 2.634**<br>(1.277)    |
| l(age^2)                     |                            | 0.007<br>(0.008)      | −0.135***<br>(0.046)  |
| l(age^3)                     |                            | 0.0003***<br>(0.0001) | 0.002***<br>(0.001)   |
| l(age^4)                     |                            |                       | −0.00001<br>(0.00001) |
| log(visceral_fat_g):age      | 0.068***<br>(0.016)        | 0.279***<br>(0.075)   | −0.239<br>(0.243)     |
| log(visceral_fat_g):l(age^2) |                            | −0.004***<br>(0.001)  | 0.013<br>(0.008)      |
| log(visceral_fat_g):l(age^3) |                            |                       | −0.0002**<br>(0.0001) |

# Оцінка моделей II

|                     |                      |                      |                       |
|---------------------|----------------------|----------------------|-----------------------|
| age:genderMale      | -0.179***<br>(0.021) | 0.928***<br>(0.283)  | 0.088<br>(0.724)      |
| l(age^2):genderMale |                      | -0.028***<br>(0.009) | 0.021<br>(0.039)      |
| l(age^3):genderMale |                      | 0.0002**<br>(0.0001) | -0.001<br>(0.001)     |
| l(age^4):genderMale |                      |                      | 0.00001<br>(0.00001)  |
| age:is_black        | 0.167***<br>(0.018)  | 0.391***<br>(0.080)  | 0.304<br>(0.698)      |
| l(age^2):is_black   |                      | -0.004***<br>(0.001) | -0.011<br>(0.041)     |
| l(age^3):is_black   |                      |                      | 0.0004<br>(0.001)     |
| l(age^4):is_black   |                      |                      | -0.00000<br>(0.00001) |

# Оцінка моделей III

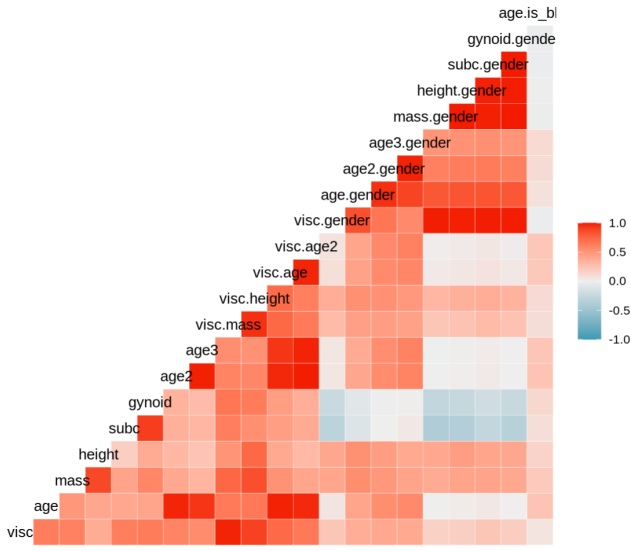
|                         |        |        |        |
|-------------------------|--------|--------|--------|
| Observations            | 16,788 | 16,788 | 16,788 |
| Adjusted R <sup>2</sup> | 0.374  | 0.381  | 0.382  |

*Note:*

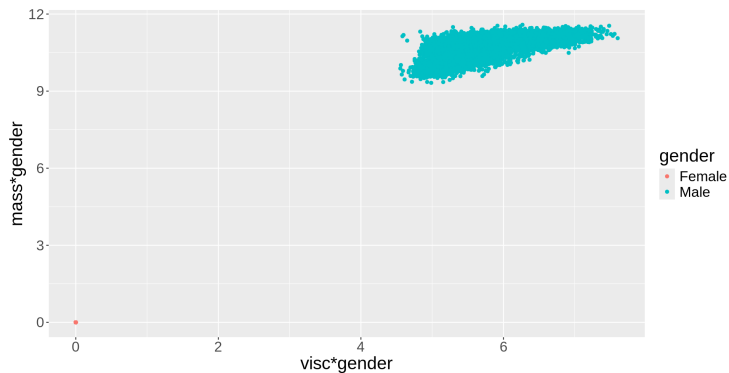
\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

# Фінальна модель. Кореляції між регресорами.

- Як фінальну модель обрано попередню модель (2) з поліномами віку.



# Фінальна модель. Кореляції між регресорами.



- Вирішено змінні з високими кореляціями не прибрати.



# Фінальна модель. Коефіцієнт при вісцеральному жиру

- Маємо наступну оцінку фінальної моделі:

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

---

|                                      |                      |
|--------------------------------------|----------------------|
| log(visceral_fat_g)                  | 0.464<br>(6.488)     |
| log(visceral_fat_g):log(lean_mass_g) | 1.043<br>(0.890)     |
| log(visceral_fat_g):height_cm        | -0.080***<br>(0.023) |
| log(visceral_fat_g):age              | 0.279***<br>(0.075)  |
| log(visceral_fat_g):l(age^2)         | -0.004***<br>(0.001) |
| log(visceral_fat_g):genderMale       | 2.051***<br>(0.721)  |

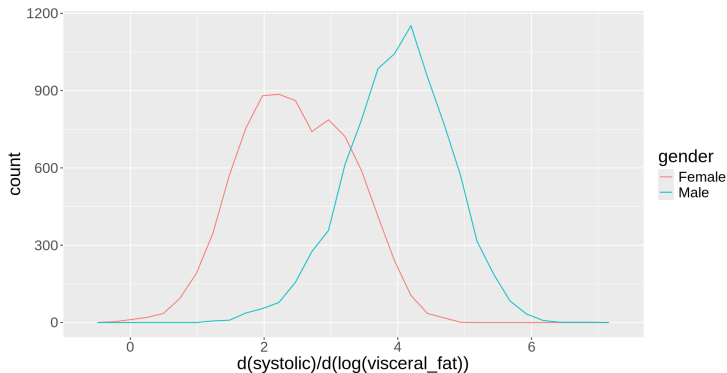
---

Note:

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

# Фінальна модель. Коефіцієнт при вісцеральному жиру

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



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

## Фінальна модель. Коефіцієнт при вісцеральному жиру

|        | $\text{median}(\beta_{\ln(\text{visceral fat})})$ |
|--------|---|
| Female | 2.45  |
| Male   | 4.04  |

- Для жінок збільшення маси вісцерального жиру на 10% в середньому підвищує систолічний тиск на 0.245 mmHg.
- Для чоловіків збільшення маси вісцерального жиру на 10% в середньому підвищує систолічний тиск на 0.404 mmHg.
- Але середньоквадратичне відхилення систолічного тиску  $\approx 15$  mmHg, тому вплив доволі малий.

# Фінальна модель. Коефіцієнт при вісцеральному жиру

---

|  |                      |
|--|----------------------|
| $\log(\text{visceral\_fat\_g})$                            | 0.464<br>(6.488)     |
| $\log(\text{visceral\_fat\_g}):\log(\text{lean\_mass\_g})$ | 1.043<br>(0.890)     |
| $\log(\text{visceral\_fat\_g}):\text{height\_cm}$          | -0.080***<br>(0.023) |
| $\log(\text{visceral\_fat\_g}):\text{age}$                 | 0.279***<br>(0.075)  |
| $\log(\text{visceral\_fat\_g}):I(\text{age}^2)$            | -0.004***<br>(0.001) |
| $\log(\text{visceral\_fat\_g}):\text{genderMale}$          | 2.051***<br>(0.721)  |

---

*Note:*

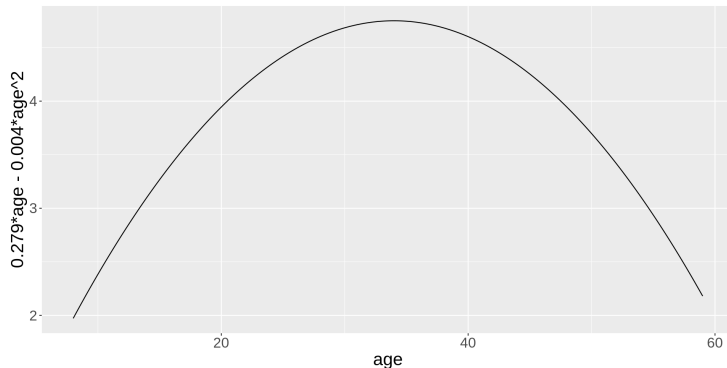
\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

# Фінальна модель. Коефіцієнт при вісцеральному жиру

- Коефіцієнт при нежировій масі не є статистично значущим.
- Коефіцієнт при зросту людини є статистично значущим та від'ємним.  
Тобто при збільшенні вісцерального жиру підвищення систолічного тиску є більшим для людей меншого зросту.
- 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.

# Фінальна модель. Коефіцієнт при вісцеральному жиру

- 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, при чому нелінійним чином:



- 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  $R_{adj.}^2 = 0.381$  виявилось малим, можуть бути невраховані важливі змінні які корелюють з масою вісцерального жиру.