

ЗВІТ

з лабораторної роботи №4
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

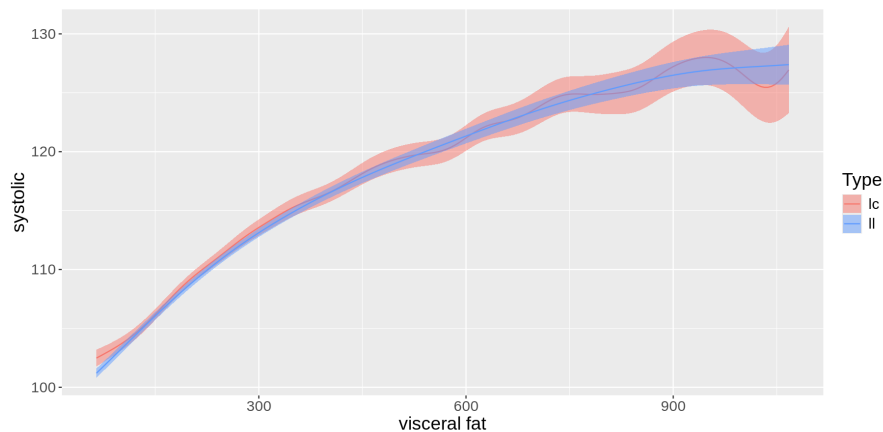
Варто зауважити, що для оцінювання непараметричних моделей початкова вибірка була розділена на тренувальну та тестову вибірки. Ширини вікон оцінювалися на тренувальній вибірці, а прогнози робилися на тестовій вибірці. Розмір обох вибірок склав 50% від розміру початкової вибірки.

2.1 Модель 1

Спочатку розглянемо модель тільки з кількістю вісцерального жиру в непараметричній частині:

$$y = m(\text{visceral fat}) + \varepsilon$$

Маємо наступні результати для локально-сталого та локально-лінійного оцінок:



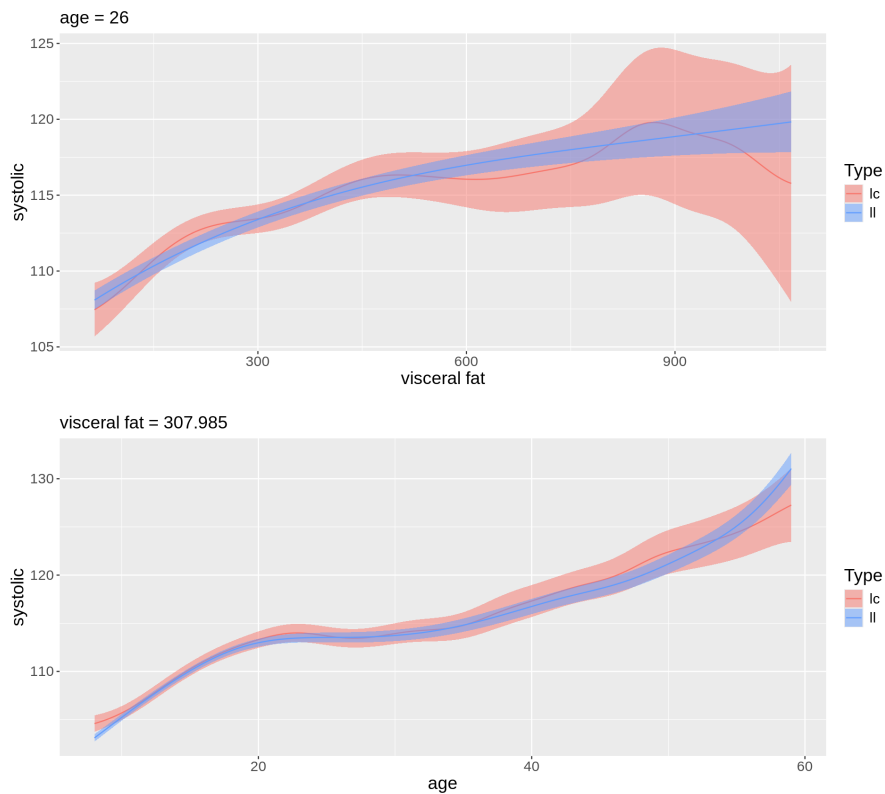
- Зі збільшенням кількості вісцерального жиру систолічний тиск помітно підвищується.

2.2 Модель 2

Додамо ще вік до непараметричної частини:

$$y = m(\text{visceral fat}, \text{age}) + \varepsilon$$

Зафіксувавши іншу змінну на медіанному рівні, маємо наступні результати для локально-сталого та локально-лінійного оцінок:



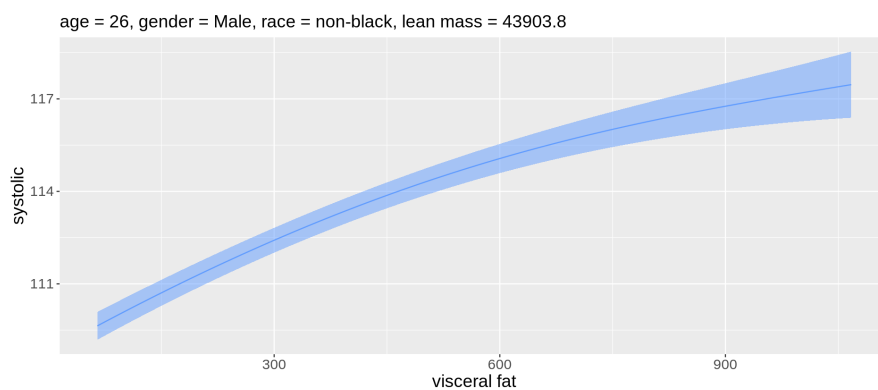
- Зі збільшенням кількості вісцерального жиру систолічний тиск підвищується, але вже менш помітно.
- Тепер підвищення систолічного тиску більш асоційовано зі збільшенням віку.

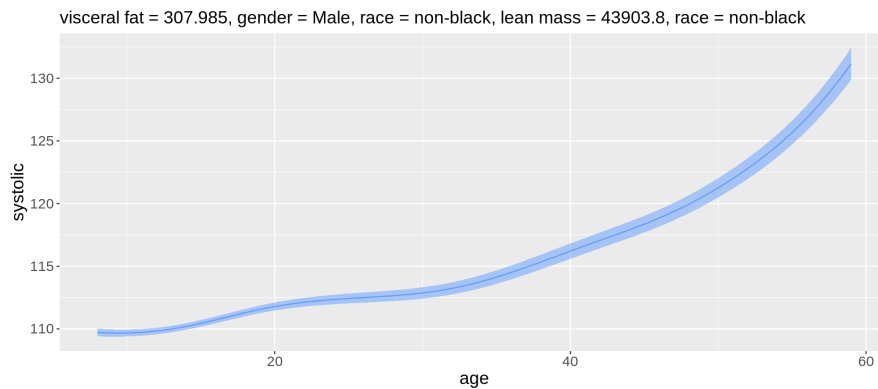
2.3 Частково лінійна модель

Розглянемо частково-лінійну модель, контролюючи стать, расу та кількість нежирової маси:

$$y = m(visceral\ fat, age) + \beta_{is\ male} \cdot is\ male + \beta_{is\ black} \cdot is\ black + \beta_{lean\ mass} \cdot \ln(lean\ mass) + \varepsilon$$

Зафіксувавши інші змінні на медіанному рівні, маємо наступні результати для локально-лінійної оцінки:





- Зі збільшенням кількості вісцерального жиру систолічний тиск підвищується, але вплив став ще меншим.
- Підвищення систолічного тиску помітно зі збільшенням віку.

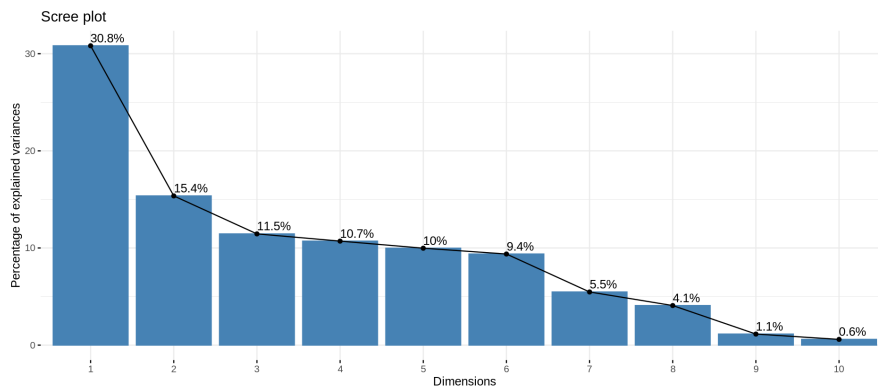
Порівняємо коефіцієнти з лінійної частини з коефіцієнтами з лінійної моделі:

	(1)	(2)
(Intercept)	-4.557 (5.193)	
genderMale	1.471*** (0.219)	1.475*** (0.337)
raceNon-Hispanic Black	3.611*** (0.233)	3.746*** (0.328)
log(lean_mass_g)	10.582*** (0.491)	10.752*** (0.775)
log(visceral_fat_g)	-1.162+ (0.675)	
age	-1.132*** (0.270)	
I(age^2)	0.006 (0.005)	
I(age^3)	0.000*** (0.000)	
log(visceral_fat_g) × age	0.332*** (0.053)	
log(visceral_fat_g) × I(age^2)	-0.005*** (0.001)	
Num.Obs.	16788	8394

Вони дуже схожі, стандартна похибка виявилась вищою через менший розмір тренувальної вибірки.

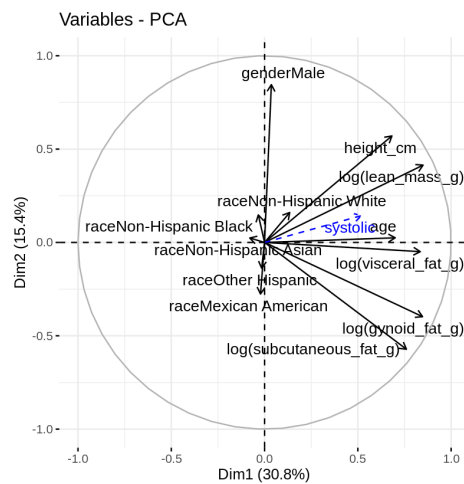
2.4 PCA

Для PCA було взято 12 змінних. Отримано наступний результат:



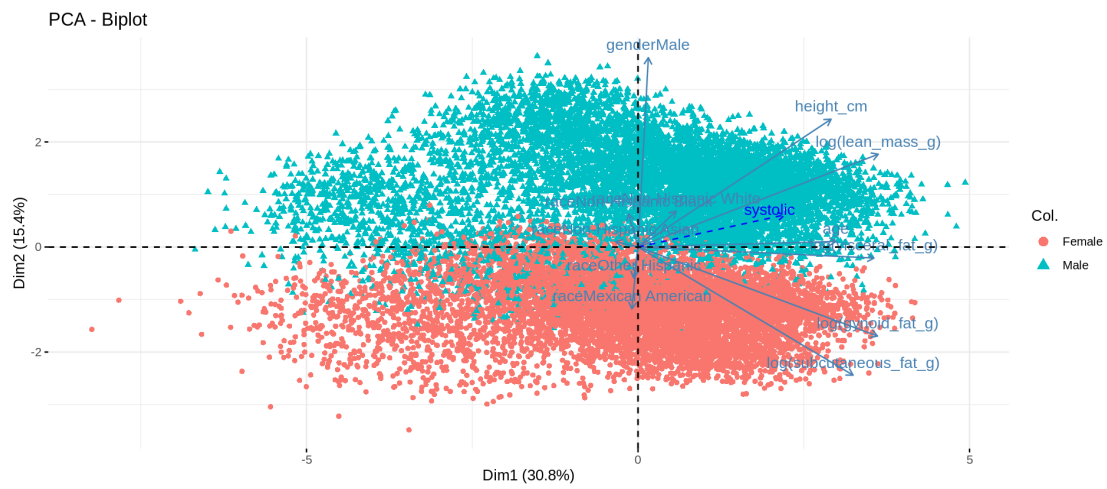
Перші 6 компонент пояснюють 87.8% дисперсії. Значного зменшення розмірності непомітно.

Також отримані наступні графіки проєкцій змінних на створені змінні.

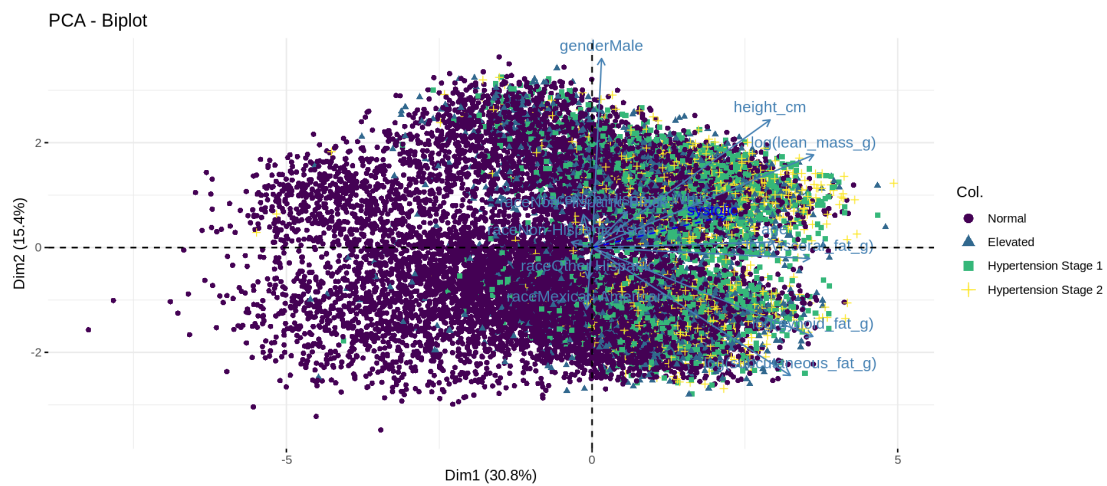


Можна припустити, що перша компонента відповідає розміру людини.

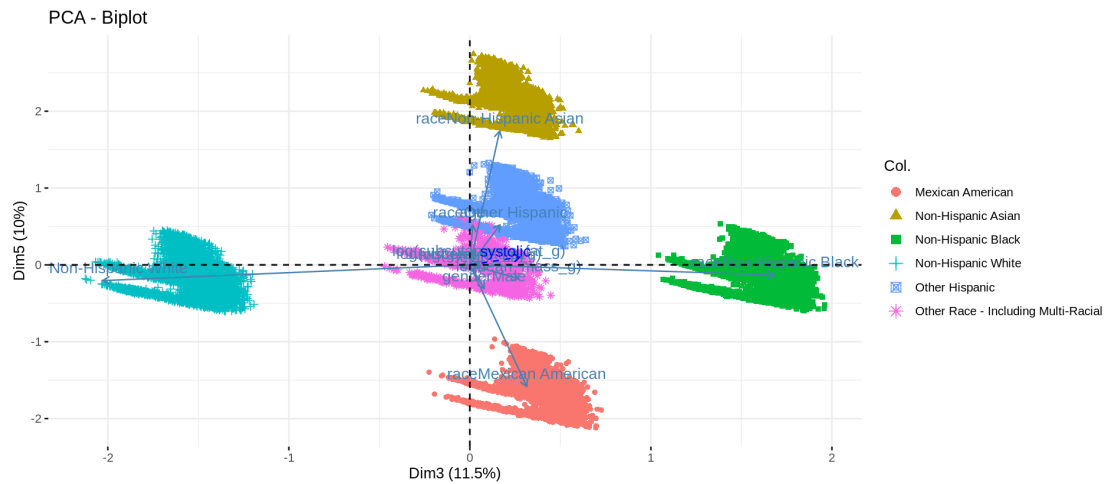
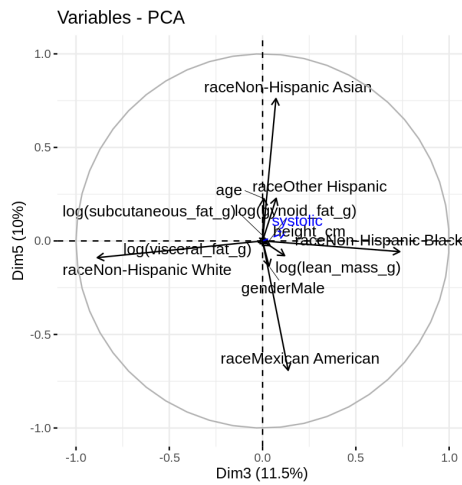
А друга компонента відповідає статі:



Але перші дві компоненти не дуже добре розрізняють артеріальний тиск:



Третя та п'ята компоненти розрізняють расу:



3 Висновки

- Непараметричні моделі дали дещо схожий результат з попередньою лабораторною роботою: зі збільшенням кількості вісцерального жиру підвищується артеріальний тиск, але сам вплив незначний. Але побудовані непараметричні моделі не враховували деякі важливі фактори взаємодії (наприклад, між кількістю жиру та статтю).
- Непараметричні моделі дали змогу побачити нелінійні зв'язки (такі як між віком та тиском).
- Значного зменшення розмірності досягнути за допомогою PCA не вдалось, але можна використати деякі компоненти як змінні які характеризують певні характеристики людини (такі як її розмір).

4 Використані джерела

[1] Лекції 11, 12 з дисципліни "Аналіз даних"