

# Лабораторна робота 2

## Статистичне виведення

Піковець Артем КМ-22

# Дослідницькі питання

- ❶ Як вік, стать та раса впливають на артеріальний тиск людини?
  - Чи погіршується артеріальний тиск зі зростанням віку?
  - Чи існують відмінності за статтю або расою?
- ❷ Як кількість та розподіл жиру в тілі людини впливає на її артеріальний тиск?
  - Чи підвищується артеріальний тиск з збільшенням кількості жиру?
  - Чи має кількість жирової маси більший вплив на артеріальний тиск ніж кількість нежирової маси?
  - Чи жир в андроїдній області має більший вплив на артеріальний тиск ніж жир в гіноїдній області?
  - Чи кількість саме вісцерального жиру найбільш пов'язана з артеріальним тиском?

- Дані були взяті з NHANES за роки 2011-2018.
- Після обробки змінних артеріального тиску (systolic та diastolic), була отримана вибірка з 28097 людей.
- Для кожного дослідницького питання були взяті змінні які його стосуються. Рядки з пропущеними даними було видалено.
- Отримано дві вибірки:

whole_data (28097 рядків)	fat_data (18436 рядків)
systolic diastolic blood_pressure_category age gender race	systolic diastolic blood_pressure_category age gender race android_fat_g android_non_fat_g gynoid_fat_g gynoid_non_fat_g visceral_fat_g subcutaneous_fat_g

- 1 Розподіл артеріального тиску
- 2 Вплив віку, статі та раси на артеріальний тиск
- 3 Вплив кількості та розподілу жиру на артеріальний тиск
- 4 Множинне тестування гіпотез

- 1 Розподіл артеріального тиску
- 2 Вплив віку, статі та раси на артеріальний тиск
- 3 Вплив кількості та розподілу жиру на артеріальний тиск
- 4 Множинне тестування гіпотез

# Дескриптивні статистики

- $\mu_X$  - сподівання
- $\sigma_X$  - середньо-квадратичне відхилення

	Довірчі інтервали (95%)	
	$\mu_X$	$\sigma_X$
systolic	[119.29, 119.73]	[18.58, 19.02]
diastolic	[67.22, 67.53]	[12.95, 13.20]

- $P_X(X = x)$  - функція ймовірностей для дискретної випадкової величини

Blood pressure category	Довірчі інтервали (95%)
	$P_X(X = x)$
Normal	[0.548, 0.560]
Elevated	[0.134, 0.142]
Hypertension Stage 1	[0.156, 0.165]
Hypertension Stage 2	[0.143, 0.151]

- В даній вибірці більше ніж 50% людей мають нормальний тиск.

# Кореляція між систолічним та діастолічним тиском

- $r_s$  - коефіцієнт кореляції Спірмана
- Побудова студентизованого та ВСа довірчих інтервалів є ресурсозатратною.
- Тому було використано невелику бутстреп-вибірку.  
 $R = 200, R_{sd} = 50$ .

Довірчі інтервали (95%) $r_s(\text{systolic}, \text{diastolic})$	
norm	[0.4905, 0.5095]
basic	[0.4902, 0.5101]
stud	[0.4894, 0.5113]
perc	[0.4916, 0.5115]
BCa	[0.4908, 0.5094]

- Присутня позитивна кореляція але недостатньо велика щоб вважати систолічний та діастолічний тиск подібними величинами.
- Надалі обчислюватись будуть тільки norm, basic та perc довірчі інтервали але з більшою бутстреп-вибіркою.

# Статистичне виведення

- 1 Розподіл артеріального тиску
- 2 Вплив віку, статі та раси на артеріальний тиск**
- 3 Вплив кількості та розподілу жиру на артеріальний тиск
- 4 Множинне тестування гіпотез



# Сподівання та дисперсії тиску серед різних вікових груп

age	Довірчі інтервали (95%)			
	systolic		diastolic	
	$\mu_X$	$\sigma_X$	$\mu_X$	$\sigma_X$
[5, 15]	[103.7, 104.3]	[9.4, 9.8]	[55.7, 56.3]	[10.9, 11.4]
(15, 25]	[111.8, 112.5]	[10.6, 11.2]	[63.3, 63.9]	[10.2, 10.8]
(25, 35]	[114.6, 115.5]	[12.5, 13.4]	[69.4, 70.1]	[10.5, 11.2]
(35, 45]	[118.7, 119.7]	[14.6, 15.7]	[73.8, 74.6]	[10.8, 11.4]
(45, 55]	[123.7, 124.9]	[16.5, 17.9]	[75.0, 75.7]	[10.5, 11.1]
(55, 65]	[129.9, 131.1]	[18.5, 19.7]	[72.7, 73.4]	[11.3, 11.9]
(65, 75]	[133.0, 134.4]	[18.5, 19.7]	[68.0, 68.9]	[11.6, 12.3]
(75, 85]	[137.9, 139.9]	[21.6, 23.2]	[63.4, 64.5]	[11.9, 12.9]

- Зі збільшенням віку систолічний тиск підвищується, та його середньоквадратичне відхилення теж.
- Діастолічний тиск спочатку підвищується, потім починає спадати. Відмінність між середньоквадратичними відхилення вже не така помітна.

# Сподівання систолічного тиску серед різних вікових груп

- $\mu_{[a,b]}$  - сподівання систолічного тиску для людей вікової групи  $[a, b]$

$$H_0 : \mu_{[a_i, b_i]} - \mu_{[a_{i+1}, b_{i+1}]} \geq 0 \quad \text{vs} \quad H_1 : \mu_{[a_i, b_i]} - \mu_{[a_{i+1}, b_{i+1}]} < 0$$

$H_0$	p	$p < 0.05$
$\mu_{[5,15]} - \mu_{(15,25]} \geq 0$	0	TRUE
$\mu_{(15,25]} - \mu_{(25,35]} \geq 0$	2.736e-26	TRUE
$\mu_{(25,35]} - \mu_{(35,45]} \geq 0$	1.305e-34	TRUE
$\mu_{(35,45]} - \mu_{(45,55]} \geq 0$	9.128e-39	TRUE
$\mu_{(45,55]} - \mu_{(55,65]} \geq 0$	5.965e-48	TRUE
$\mu_{(55,65]} - \mu_{(65,75]} \geq 0$	5.786e-11	TRUE
$\mu_{(65,75]} - \mu_{(75,85]} \geq 0$	2.252e-16	TRUE

- Усі  $p$ -значення дуже малі. Старші вікові групи в середньому мають вищий систолічний тиск.

# Дисперсії систолічного тиску серед різних вікових груп

- $\sigma_{[a,b]}$  - середньоквадратичне відхилення систолічного тиску для людей вікової групи  $[a, b]$

$$H_0 : \sigma_{[a_i, b_i]} - \sigma_{[a_{i+1}, b_{i+1}]} \geq 0 \quad \text{vs} \quad H_1 : \sigma_{[a_i, b_i]} - \sigma_{[a_{i+1}, b_{i+1}]} < 0$$

$H_0$	p	$p < 0.05$
$\sigma_{[5,15]} - \sigma_{[15,25]} \geq 0$	4.8e-13	TRUE
$\sigma_{[15,25]} - \sigma_{[25,35]} \geq 0$	7.859e-15	TRUE
$\sigma_{[25,35]} - \sigma_{[35,45]} \geq 0$	2.801e-10	TRUE
$\sigma_{[35,45]} - \sigma_{[45,55]} \geq 0$	1.256e-06	TRUE
$\sigma_{[45,55]} - \sigma_{[55,65]} \geq 0$	1.077e-05	TRUE
$\sigma_{[55,65]} - \sigma_{[65,75]} \geq 0$	0.5026	FALSE
$\sigma_{[65,75]} - \sigma_{[75,85]} \geq 0$	7.21e-11	TRUE

- Майже усі  $p$ -значення є малими. Старші вікові групи мають більший розкид систолічного тиску.

# Кореляція між віком та артеріальним тиском

- $R$  - кількість бутстреп-вибірок для підрахунку довірчих інтервалів

		Percentile CI (95%), $R = 1000$
	$r_s(\text{age}, \text{systolic})$	[0.601, 0.616]
$\text{age} < 50$	$r_s(\text{age}, \text{diastolic})$	[0.568, 0.589]
$\text{age} \geq 50$	$r_s(\text{age}, \text{diastolic})$	[-0.355, -0.320]

- Зі збільшенням віку систолічний тиск теж збільшується; діастолічний тиск теж спочатку збільшується, але потім зменшується.

# Розподіл віку серед людей з нормальним тиском та людей з гіпертонією

- $Q_X(q) = F_X^{-1}(q)$  - квантиль
- $R$  - кількість бутстреп-вибірок для підрахунку стандартної похибки вибіркових квантилів

Довірчі інтервали (95%), $R = 300$		
	age	
Blood pressure	$Q(0.25)$	$Q(0.75)$
Normal	[13.04, 14.96]	[41.15, 42.85]
Hypertension	[43.04, 44.96]	[68.15, 69.85]

- Люди з нормальним тиском значно молодші за людей з гіпертонією.

$$H_0 : Q_{normal}(0.75) - Q_{hypertension}(0.25) \geq 0 \quad \text{vs} \quad H_1 : Q_{normal}(0.75) - Q_{hypertension}(0.25) < 0$$
$$p = 0.001358 < 0.05$$

# Відмінність артеріального тиску за статтю

Довірчі інтервали (95%)		
	$\mu_{systolic}$	$\mu_{diastolic}$
Female	[117.73, 118.38]	[66.43, 66.83]
Male	[120.72, 121.32]	[67.92, 68.38]

- Чоловіки у середньому мають вищий артеріальний тиск ніж жінки.

$$H_0 : \mu^{female} - \mu^{male} \geq 0 \quad \text{vs} \quad H_1 : \mu^{female} - \mu^{male} < 0$$

$$p_{systolic} = 1.352e - 40 < 0.05$$

$$p_{diastolic} = 1.32e - 22 < 0.05$$

# Відмінність артеріального тиску за расою

- $M_X = F_X^{-1}(0.5)$  - медіана

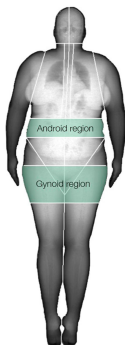
Довірчі інтервали (95%)			$R = 300$
	$\mu_{systolic}$	$\mu_{diastolic}$	$M_{age}$
Other Race	[115.3, 117.3]	[65.7, 67.2]	[23.6, 28.4]
Mexican American	[116.5, 117.5]	[65.2, 66.0]	[30.6, 33.4]
Non-Hispanic Asian	[116.9, 118.0]	[68.7, 69.5]	[36.7, 39.3]
Other Hispanic	[118.3, 119.7]	[65.9, 66.8]	[38.3, 41.7]
Non-Hispanic White	[119.6, 120.3]	[67.3, 67.8]	[42.9, 45.1]
Non-Hispanic Black	[121.9, 122.9]	[67.6, 68.3]	[36.6, 39.4]

- Спостерігається відмінність систолічного тиску між расами, але це може бути пов'язано з відмінністю розподілу віку між расами.

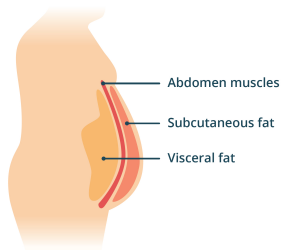
- 1 Розподіл артеріального тиску
- 2 Вплив віку, статі та раси на артеріальний тиск
- 3 Вплив кількості та розподілу жиру на артеріальний тиск**
- 4 Множинне тестування гіпотез



## Розташування андроїдної та гіноїдної області



## Розташування вісцерального та підшкірного жиру



# Розподіл жиру в андроїдній області

		Довірчі інтервали (95%), $R = 300$		
android		$Q(0.2)$	$Q(0.5)$	$Q(0.8)$
fat	Male	[655, 691]	[1584, 1650]	[2825, 2911]
	Female	[861, 904]	[1749, 1820]	[3044, 3165]
non fat	Male	[2568, 2681]	[3881, 3958]	[5057, 5155]
	Female	[2146, 2205]	[2951, 3010]	[3941, 4020]
fat percent	Male	[0.194, 0.199]	[0.299, 0.305]	[0.379, 0.384]
	Female	[0.278, 0.286]	[0.378, 0.383]	[0.449, 0.452]

В андроїдній області:

- Жінки мають більше жиру ніж чоловіки.
- Чоловіки мають більше нежирової маси ніж жінки.
- У середньому, маса жиру становить приблизно 30% для чоловіків та 38% для жінок від загальної маси області.

# Розподіл жиру в гіноїдній області

		Довірчі інтервали (95%), $R = 300$		
gynoid		Q(0.2)	Q(0.5)	Q(0.8)
fat	Male	[1971, 2041]	[3212, 3292]	[4614, 4741]
	Female	[2886, 2980]	[4325, 4428]	[6198, 6348]
non fat	Male	[6052, 6349]	[8421, 8532]	[10377, 10515]
	Female	[4827, 4936]	[6257, 6347]	[7845, 7978]
fat percent	Male	[0.232, 0.236]	[0.286, 0.289]	[0.345, 0.348]
	Female	[0.363, 0.366]	[0.413, 0.416]	[0.458, 0.460]

В гіноїдній області:

- Жінки мають більше жиру ніж чоловіки.
- Чоловіки мають більше нежирової маси ніж жінки.
- У середньому, маса жиру становить приблизно 29% для чоловіків та 41% для жінок від загальної маси області.

# Розподіл жиру між андроїдною та гіноїдною областю

Довірчі інтервали (95%), $R = 300$		
	$M_{\frac{\text{android fat}}{\text{gynoid fat}}}$	$M_{\frac{\text{android non fat}}{\text{gynoid non fat}}}$
Female	[0.393, 0.400]	[0.474, 0.477]
Male	[0.463, 0.475]	[0.465, 0.468]

- Чоловіки більш схильні зберігати жир в андроїдній області ніж жінки. З нежировою масою це вже не так.

$$H_0 : M_{\frac{\text{female}}{\text{android fat}} \text{ gynoid fat}} - M_{\frac{\text{male}}{\text{android fat}} \text{ gynoid fat}} \geq 0 \quad \text{vs} \quad H_1 : M_{\frac{\text{female}}{\text{android fat}} \text{ gynoid fat}} - M_{\frac{\text{male}}{\text{android fat}} \text{ gynoid fat}} < 0$$

$$p = 8.63e - 109 < 0.05$$

$$H_0 : M_{\frac{\text{male}}{\text{android non fat}} \text{ gynoid non fat}} - M_{\frac{\text{female}}{\text{android non fat}} \text{ gynoid non fat}} \geq 0 \quad \text{vs} \quad H_1 : M_{\frac{\text{male}}{\text{android non fat}} \text{ gynoid non fat}} - M_{\frac{\text{female}}{\text{android non fat}} \text{ gynoid non fat}} < 0$$

$$p = 1.189e - 14 < 0.05$$

# Розподіл вісцерального та підшкірного жиру

		Довірчі інтервали (95%), $R = 300$		
gynoid		Q(0.2)	Q(0.5)	Q(0.8)
visceral	Female	[131, 138]	[297, 312]	[568, 592]
	Male	[193, 198]	[324, 338]	[604, 625]
subcutaneous	Female	[926, 964]	[1570, 1614]	[2381, 2438]
	Male	[347, 374]	[973, 1018]	[1661, 1708]
<u>visceral</u> <u>subcutaneous</u>	Female	[0.128, 0.131]	[0.181, 0.185]	[0.270, 0.279]
	Male	[0.268, 0.277]	[0.421, 0.432]	[0.657, 0.673]

- Чоловіки мають більше вісцерального жиру ніж жінки.
- Жінки мають більше підшкірного жиру ніж чоловіки.
- Кількість вісцерального жиру менше ніж кількість інших розглянутих видів жиру.
- Чоловіки більше схильні до накопичення вісцерального жиру ніж жінки.

# Кореляція між жировою та нежировою масою

	Percentile CI (95%), $R = 1000$	
	Female	Male
$r_s(\text{android fat, android non fat})$	[0.846, 0.860]	[0.883, 0.894]
$r_s(\text{gynoid fat, gynoid non fat})$	[0.850, 0.864]	[0.738, 0.759]

- Досить великі кореляції. Збільшення жирової маси асоційовано зі збільшенням нежирової маси.
- Яка саме з цих змінних буде мати вплив на артеріальний тиск?

# Кореляція між різними видами жиру

	Percentile CI (95%), $R = 1000$	
	Female	Male
$r_s(\text{android fat, gynoid fat})$	[0.897, 0.907]	[0.916, 0.924]
$r_s(\text{visceral fat, subcutaneous fat})$	[0.831, 0.846]	[0.805, 0.822]

- Досить великі кореляції. Збільшення кількості одного виду жиру асоційовано зі збільшенням кількості іншого виду жиру.
- Який саме вид жиру буде мати вплив на артеріальний тиск?

# Зміна кількості жиру з віком

- На початку життя людини (0-20 років) розмір її тіла помітно збільшується, і кількість жиру теж. Чи збільшується кількість жиру і після 26 років?

$age \geq 26$	Percentile CI (95%), $R = 1000$	
	Female	Male
$r_s(\text{age, android fat})$	[0.103, 0.156]	[0.113, 0.169]
$r_s(\text{age, gynoid fat})$	[-0.0016, 0.0554]	[-0.0858, -0.0293]
$r_s(\text{age, visceral fat})$	[0.290, 0.341]	[0.365, 0.413]
$r_s(\text{age, subcutaneous fat})$	[0.0265, 0.0819]	[0.0239, 0.0815]

- Для дорослих людей, зі збільшенням віку найбільш помітно збільшується тільки вісцеральний жир на відмінну від гіноїдного та підшкірного.



# Кореляції жирової та нежирової маси з артеріальним тиском

		Percentile CI (95%), $R = 1000$	
android		$r_s(., systolic)$	$r_s(., diastolic)$
fat	Female	[0.454, 0.486]	[0.377, 0.412]
	Male	[0.470, 0.505]	[0.457, 0.491]
non fat	Female	[0.488, 0.520]	[0.413, 0.449]
	Male	[0.539, 0.570]	[0.509, 0.539]
fat percent	Female	[0.321, 0.356]	[0.256, 0.293]
	Male	[0.263, 0.300]	[0.269, 0.307]

- Кореляції присутні, підвищення жирової та нежирової маси асоційовано з підвищенням артеріального тиску.
- Жирова чи нежирова маса має більший вплив на артеріальний тиск?
  - Кореляція нежирової маси з артеріальним тиском більша за кореляцію жирової маси з артеріальним тиском.
  - З іншого боку, збільшення проценту саме жиру асоційовано з підвищенням артеріального тиску. Тому, мабуть, жирова маса має більший вплив на артеріальний тиск.

# Кореляції жирової та нежирової маси з артеріальним тиском

		Percentile CI (95%), $R = 1000$	
gynoid		$r_s(., \text{ systolic})$	$r_s(., \text{ diastolic})$
fat	Female	[0.392, 0.426]	[0.340, 0.377]
	Male	[0.386, 0.423]	[0.337, 0.375]
non fat	Female	[0.403, 0.440]	[0.328, 0.368]
	Male	[0.480, 0.513]	[0.427, 0.460]
fat percent	Female	[0.192, 0.23]	[0.178, 0.217]
	Male	[-0.037, 0.003]	[-0.049, -0.007]

- Кореляції присутні, але менші ніж в андрійдній області.
- Для чоловіків кореляція нежирової маси з артеріальним тиском більша за кореляцію жирової маси з артеріальним тиском. Для жінок вони майже однакові.
- Збільшення проценту жиру асоційовано з підвищенням артеріального тиску тільки для жінок.

# Вплив розподілу маси між андроїдною та гіноїдною областю

	Percentile CI (95%), $R = 1000$	
	Female	Male
$r_s\left(\frac{\text{android fat}}{\text{gynoid fat}}, \text{systolic}\right)$	[0.389, 0.423]	[0.476, 0.509]
$r_s\left(\frac{\text{android non fat}}{\text{gynoid non fat}}, \text{systolic}\right)$	[0.344, 0.381]	[0.285, 0.323]
$r_s\left(\frac{\text{android fat}}{\text{gynoid fat}}, \text{diastolic}\right)$	[0.302, 0.338]	[0.514, 0.543]
$r_s\left(\frac{\text{android non fat}}{\text{gynoid non fat}}, \text{diastolic}\right)$	[0.306, 0.342]	[0.327, 0.363]

- Кореляції позитивні і достатньо помітні.
- Більше зосередження жирової та нежирової маси в андроїдній області свідчить про вищий артеріальний тиск.
- Для чоловіків кореляції вищі для відношень жирових мас ніж нежирових.

# Кореляції вісцерального та підшкірного жиру з артеріальним тиском

	Percentile CI (95%), $R = 1000$	
	Female	Male
$r_s(\text{visceral fat, systolic})$	[0.476, 0.508]	[0.476, 0.508]
$r_s(\text{subcutaneous fat, systolic})$	[0.404, 0.438]	[0.400, 0.434]
$r_s(\text{visceral fat, diastolic})$	[0.409, 0.445]	[0.518, 0.548]
$r_s(\text{subcutaneous fat, diastolic})$	[0.322, 0.358]	[0.368, 0.406]

- Присутні позитивні кореляції.
- Вісцеральний жир має більшу кореляцію з артеріальним тиском ніж підшкірний жир.

# Статистичне виведення

- 1 Розподіл артеріального тиску
- 2 Вплив віку, статі та раси на артеріальний тиск
- 3 Вплив кількості та розподілу жиру на артеріальний тиск
- 4 Множинне тестування гіпотез

# Множинне тестування гіпотез I

- Використаємо метод Беньяміні-Хохберга для усіх протестованих гіпотез.

$H_0$	$p_{BH}$	$p_{BH} < 0.05$
$\mu_{[5,15]} - \mu_{(15,25]} \geq 0$	0	TRUE
$\mu_{(15,25]} - \mu_{(25,35]} \geq 0$	7.426e-26	TRUE
$\mu_{(25,35]} - \mu_{(35,45]} \geq 0$	4.131e-34	TRUE
$\mu_{(35,45]} - \mu_{(45,55]} \geq 0$	3.469e-38	TRUE
$\mu_{(45,55]} - \mu_{(55,65]} \geq 0$	3.778e-47	TRUE
$\mu_{(55,65]} - \mu_{(65,75]} \geq 0$	8.457e-11	TRUE
$\mu_{(65,75]} - \mu_{(75,85]} \geq 0$	4.754e-16	TRUE
$\sigma_{[5,15]} - \sigma_{(15,25]} \geq 0$	7.6e-13	TRUE
$\sigma_{(15,25]} - \sigma_{(25,35]} \geq 0$	1.493e-14	TRUE
$\sigma_{(25,35]} - \sigma_{(35,45]} \geq 0$	3.548e-10	TRUE
$\sigma_{(35,45]} - \sigma_{(45,55]} \geq 0$	1.491e-06	TRUE

# Множинне тестування гіпотез II

$\sigma_{(45,55]} - \sigma_{(55,65]} \geq 0$	1.204e-05	TRUE
$\sigma_{(55,65]} - \sigma_{(65,75]} \geq 0$	0.5026	FALSE
$\sigma_{(65,75]} - \sigma_{(75,85]} \geq 0$	9.786e-11	TRUE
$Q_{normal}(0.75) - Q_{hypertension}(0.25) \geq 0$	0.001433	TRUE
$\mu_{systolic}^{female} - \mu_{systolic}^{male} \geq 0$	6.423e-40	TRUE
$\mu_{diastolic}^{female} - \mu_{diastolic}^{male} \geq 0$	3.134e-22	TRUE
$M_{\frac{android\ fat}{gynoid\ fat}}^{female} - M_{\frac{android\ fat}{gynoid\ fat}}^{male} \geq 0$	8.199e-108	TRUE
$M_{\frac{android\ non\ fat}{gynoid\ non\ fat}}^{male} - M_{\frac{android\ non\ fat}{gynoid\ non\ fat}}^{female} \geq 0$	2.055e-14	TRUE

---

- Помітний вплив віку на артеріальний тиск. Зі зростанням віку систолічний тиск збільшується, а діастолічний тиск спочатку збільшується і потім зменшується.
- Спостерігається відмінність артеріального тиску за статтю: чоловіки у середньому мають вищий артеріальний тиск ніж жінки.
- Розподіл жиру та його вплив на артеріальний тиск відрізняється між чоловіками та жінками.
- Збільшення жирової маси асоційовано з підвищенням артеріального тиску.
- Більше скупчення жиру в андроїдній області у порівнянні з гіноїдною областю асоційовано з підвищенням артеріального тиску.
- Залишається проблемою встановити чи насправді існує вплив жирової або нежирової маси на артеріальний тиск, та порівняти вплив різних видів жиру.