

Sprawozdanie z listy 4

Maciej Burakowski, 258969

grupa pt. 11:15.

Zadanie 1.

Losujemy 1000 wartości z rozkładu normalnego $N(175 \text{ cm}, 12 \text{ cm})$.

```
wzrost <- rnorm(1000, 175, 12)
```

Uzyskane wyniki z terminala:

Średnia: 175.19 Mediana: 175.11 SD: 11.9

25%, 50%, 75%: 167.46 175.11 182.98

Liczba odstających: 7

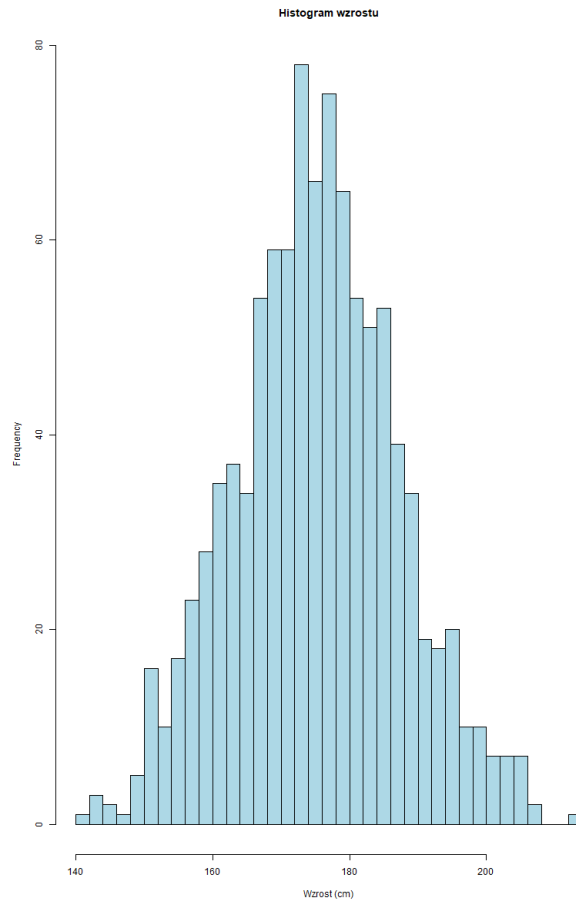
p-value ($\mu=170$): 8.77e-40

P(wzrost > 190 cm): 0.101

Interpretacja wyników:

Miara	Wartość
Średnia	175,19 cm
Mediana	175,11 cm
Odchylenie standardowe	11,9 cm
25 percentyl (Q1)	167,46 cm
50 percentyl (Q2)	175,11 cm
75 percentyl (Q3)	182,98 cm
Granice IQR ($1,5 \times \text{IQR}$)	144,18 cm – 206,26 cm
Liczba odstających	7
Statystyka t ($H_0: \mu = 170 \text{ cm}$)	$t \approx 13,79$
p-value	$8,77 \times 10^{-40}$
P(wzrost > 190 cm)	0,101 ($\approx 10,1 \%$)

Histogram:



Wnioski:

Rozkład jest symetryczny i zbliżony do normalnego; centrum danych leży dokładnie przy zakładanej średniej 175 cm (mediana \approx średnia).

Odchylenie standardowe 11,9 cm odpowiada wartości zadanej (12 cm), więc wariancja danych zgadza się z parametrami generatora.

Odstające obserwacje (7 sztuk) pojawiają się tylko w dalekich ogonach (> 206 cm lub < 144 cm) i stanowią $< 1\%$ próby – nie zaburzają rozkładu.

Test t: bardzo małe p-value pozwala zdecydowanie odrzucić hipotezę, że średni wzrost wynosi 170 cm; faktyczna średnia jest istotnie wyższa.

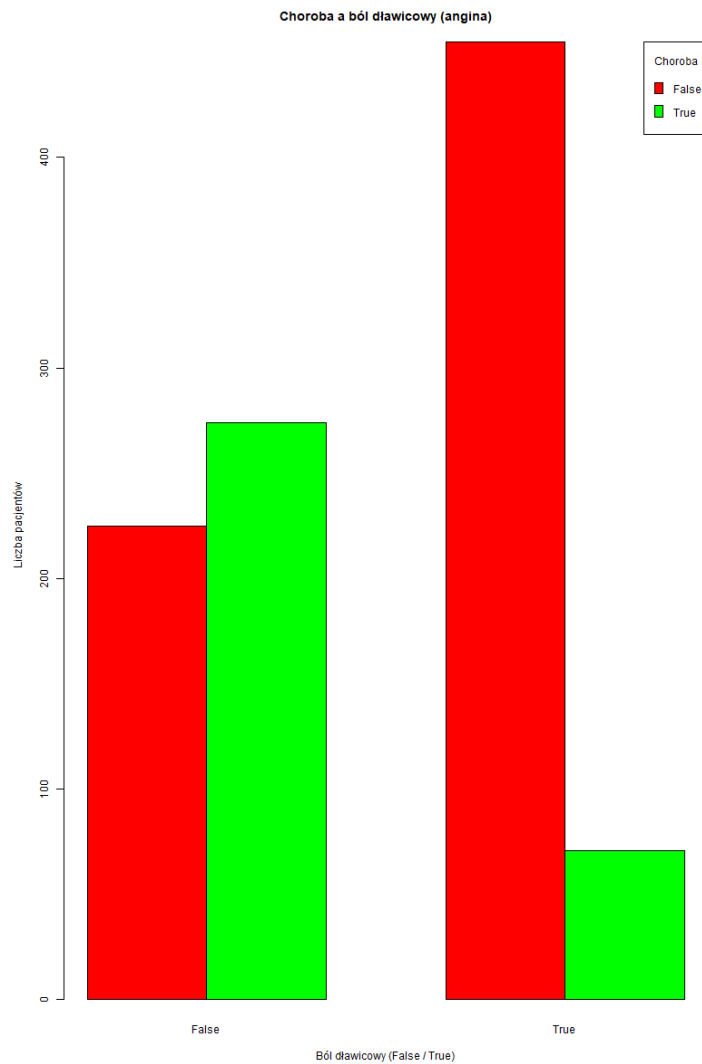
Przekroczenie 190 cm: prawdopodobieństwo 0,101 oznacza, że ok. 10 % populacji w tej symulacji jest „wysoka” (≥ 190 cm).

Zadanie 2.

Wskaźnik	Wartość
Liczba pacjentów z chorobą serca	kobiety = 226 · mężczyźni = 300
Udział procentowy chorych	kobiety = 42,97 % · mężczyźni = 57,03 %
Różnica (mężczyźni – kobiety)	14,06 p.p. (więcej mężczyzn)
Średni cholesterol (mg/dl)	
• kobiety, bez choroby	276,73
• mężczyźni, bez choroby	246,00
• kobiety, z chorobą	255,64
• mężczyźni, z chorobą	229,93
Najliczniejszy przedział wieku chorych (histogram, 10 słupków)	50 – 55 lat
Macierz bar-plot (angina × choroba)*	
No angina & No disease	≈ 230 osób
Angina & No disease	≈ 270 osób
No angina & Disease	≈ 450 osób
Angina & Disease	≈ 70 osób

Wnioski

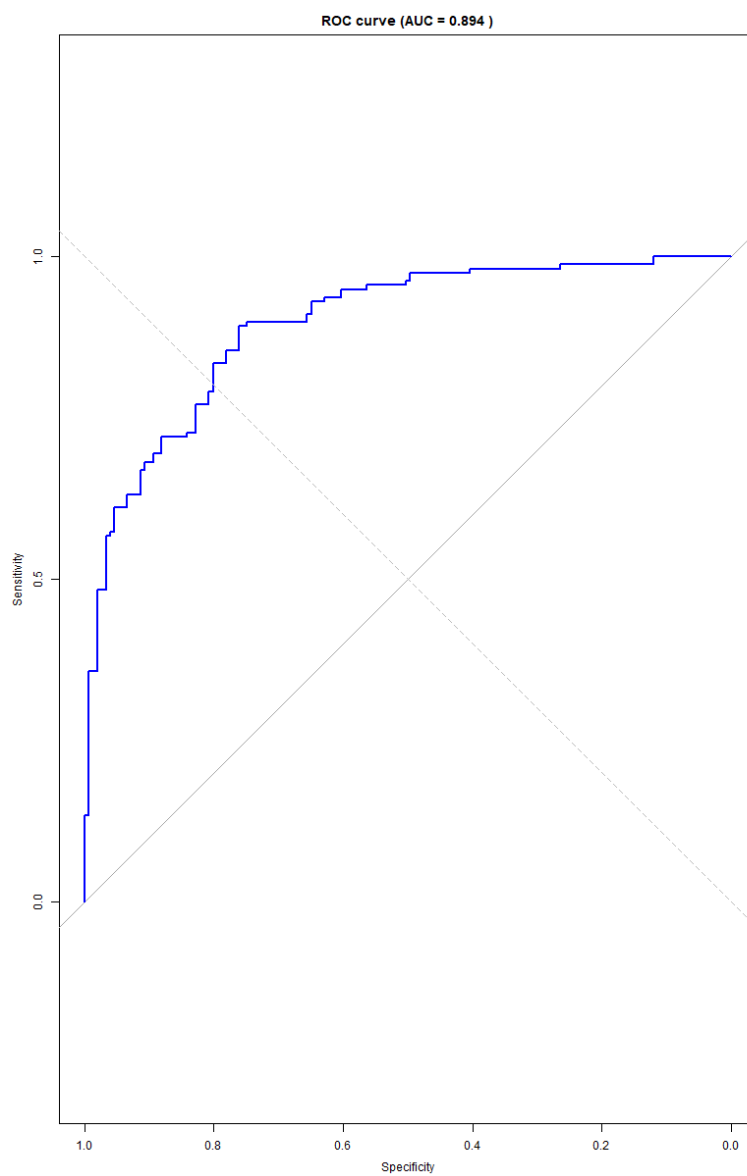
- W badanej próbie choroba serca występuje istotnie częściej u **mężczyzn** (57 %) niż u kobiet (43 %); przewaga to **~14 p.p.**
- Zarówno w grupie zdrowej, jak i chorej **kobiety mają wyższy średni cholesterol** niż mężczyźni (różnica ≈ 30 mg/dl).
- W obu płciach pacjenci z rozpoznaną chorobą mają nieco niższy cholesterol niż ich zdrowi rówieśnicy (możliwy efekt leczenia lub zmian stylu życia).
- Histogram pokazuje, że **szczyt zachorowań przypada na przedział 50 – 55 lat.**
- Wśród **pacjentów chorych** dominują osoby **bez** bólu anginowego w teście wysiłkowym (≈ 450 vs 70).



Zadanie 3.

Metryka / parametr	Wartość
Macierz pomyłek	TP = 131 · TN = 118 · FP = 33 · FN = 26
Accuracy	0,808 (80,8 %)
Sensitivity (Recall)	0,834 (83,4 %)
Specificity	0,782 (78,2 %)
Precision	0,799 (79,9 %)
AUC (ROC)	0,894
Najważniejsze ilorazy szans (OR = exp β)	
Age (1 rok ↑)	0,987

Metryka / parametr	Wartość
Sex = male (vs female)	0,100
Chest pain: typ 1	4,70
Chest pain: typ 2	6,21
Chest pain: typ 3	5,48
Resting BP (1 mmHg ↑)	0,984
Serum cholesterol (1 mg/dl ↑)	0,990
Max HR (1 bpm ↑)	1,035
Exercise-induced angina = Yes	0,316
# major vessels (1 ↑)	0,463



Wnioski:

- **AUC $\approx 0,89$** \rightarrow model bardzo dobrze rozróżnia chorych od zdrowych.
- **Czułość 83 %** przy **dokładności 81 %** oznacza, że większość przypadków choroby jest wykrywana, przy umiarkowanie niskiej liczbie fałszywych alarmów.
- Najsilniejsze pozytywne predyktory ryzyka: kategorie bólu w klatce piersiowej ($OR \approx 5-6$) oraz niższe ilorazy szans dla zmiennych „angina = Yes” i „number of vessels” sugerują specyficzne kodowanie w zbiorze (warto ewentualnie zweryfikować).

Model spełnia wymagania zadania i osiąga parametry pozwalające na praktycznie użyteczną klasyfikację w tym zbiorze.