Statystyka Opisowa z Demografią oraz Biostatystyka

Weryfikacja hipotez statystycznych

Aleksander Denisiuk

denisjuk@euh-e.edu.pl

Elbląska Uczelnia Humanistyczno-Ekonomiczna

ul. Lotnicza 2

82-300 Elbląg

Weryfikacja hipotez statystycznych

Najnowsza wersja tego dokumentu dostępna jest pod adresem

`http://denisjuk.euh-e.edu.pl/`

Oszacowanie parametrów rozkładu normalnego

- Mamy do czynienia z losową próbką wielkości N zmiennej o rozkładzie normalnym
  - Oszacowanie wartości średniej

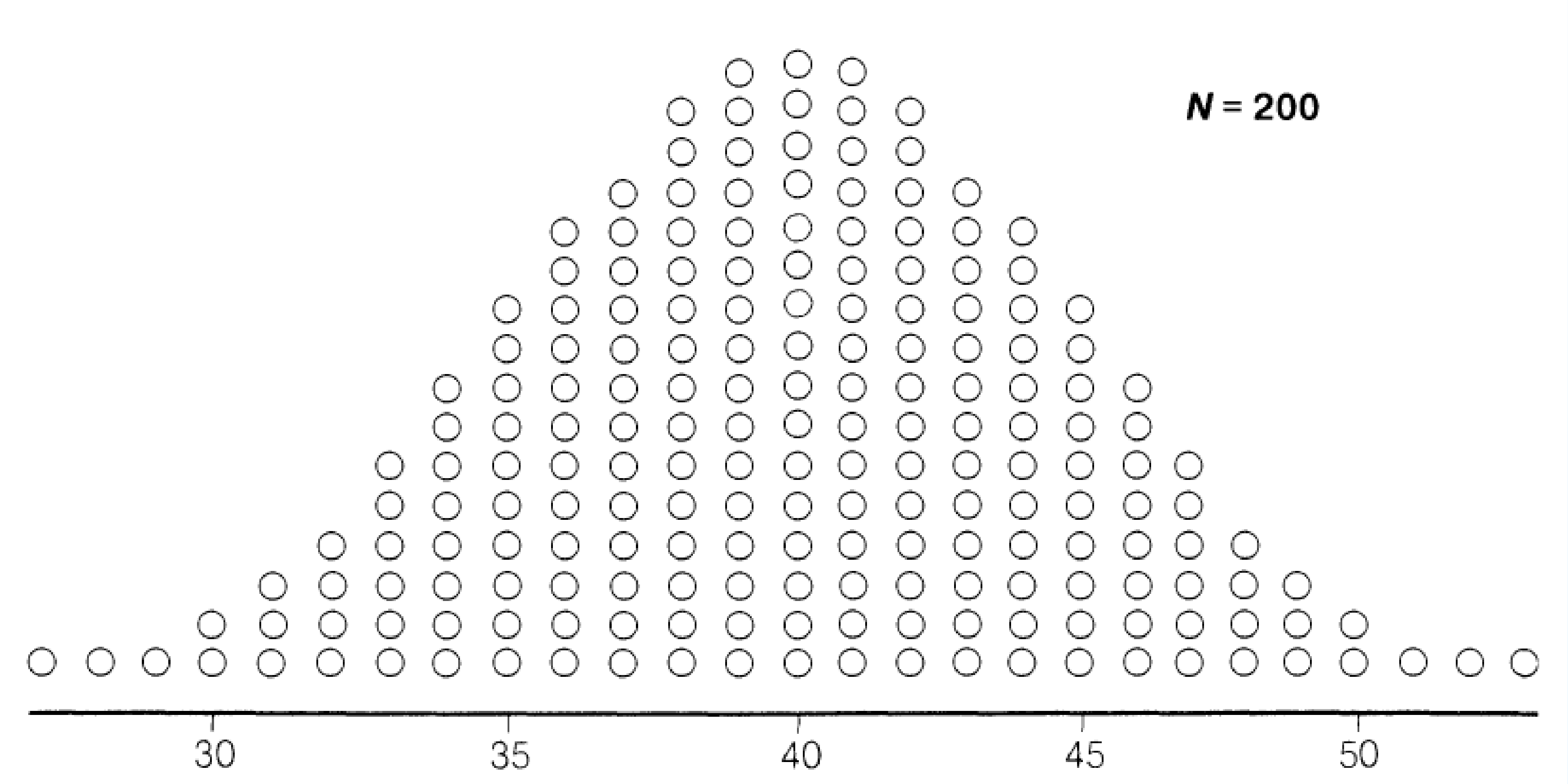
$$\bar{X} = \frac{\sum X_i}{N}$$

  - Oszacowanie odchylenia standardowego

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{N - 1}$$

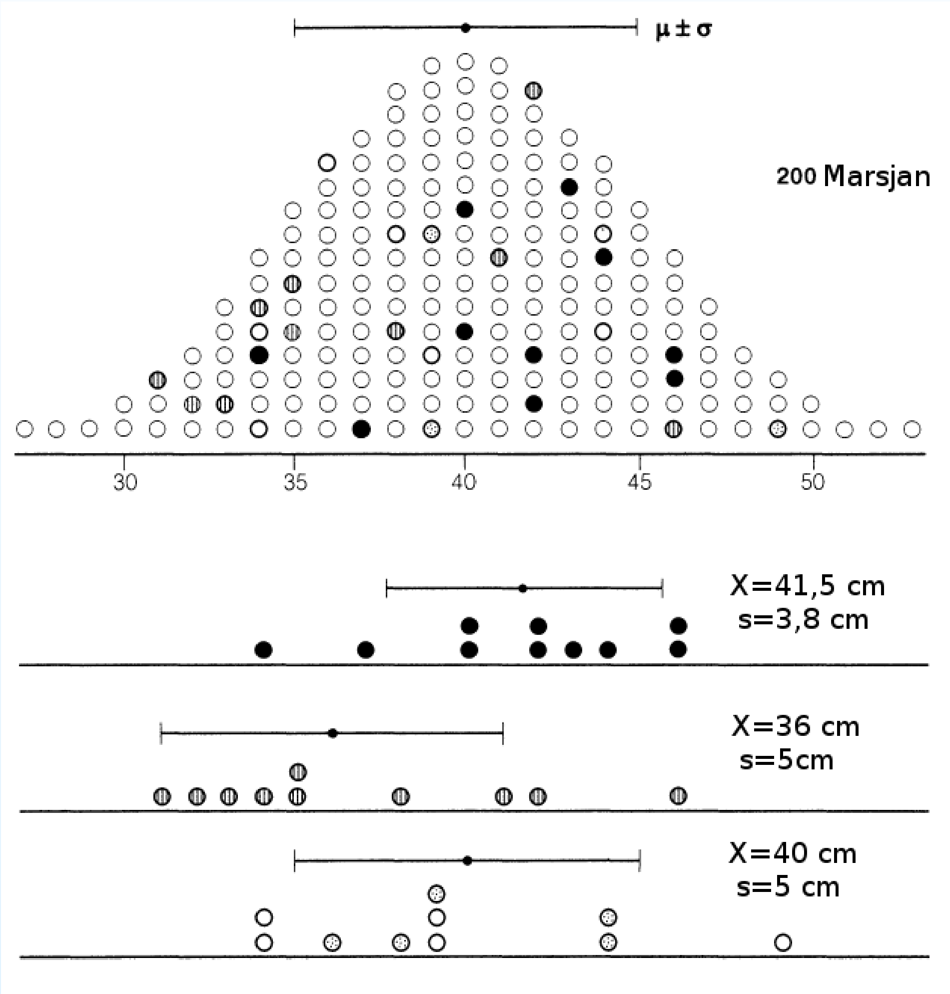
Standardowy błąd oceny wartości średniej

- Dany jest rozkład normalny $N(40, 5)$ (wzrost Marsjan, źródło: Stanton A. Glantz, *Primer of Biostatistics*)



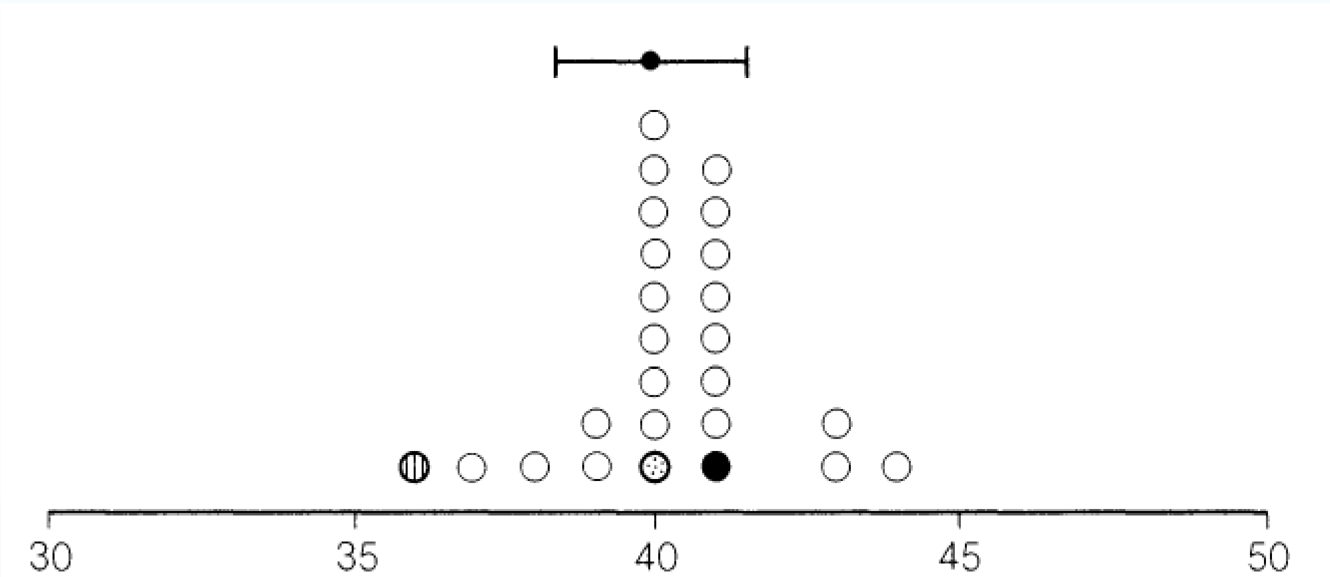
Oceny wartości średniej

- Oceniamy wartość średnią na podstawie losowej próbki 10 Marsjan



Jeszcze 25 próbek

- Ogólem jest ponad 10^{16} próbek z 10 Marsjan



- Rozkład średnich z próbki zbliża się do normalnego dla dowolnego rozkładu
  - Dla normalnego jest zawsze normalnym
- Średnia zgadza się ze średnią rozkładu początkowego

Błąd standardowy

- Odchylenie standardowe $\bar{X}_i$ jest mniejsze od odhcylenie standardowego X
- $\sigma_{\bar{X}}$ jest *błędem standardowym* wartości średniej z próby

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

- $s_{\bar{X}}$ jest oceną błędu standardowego wartości średniej z próby

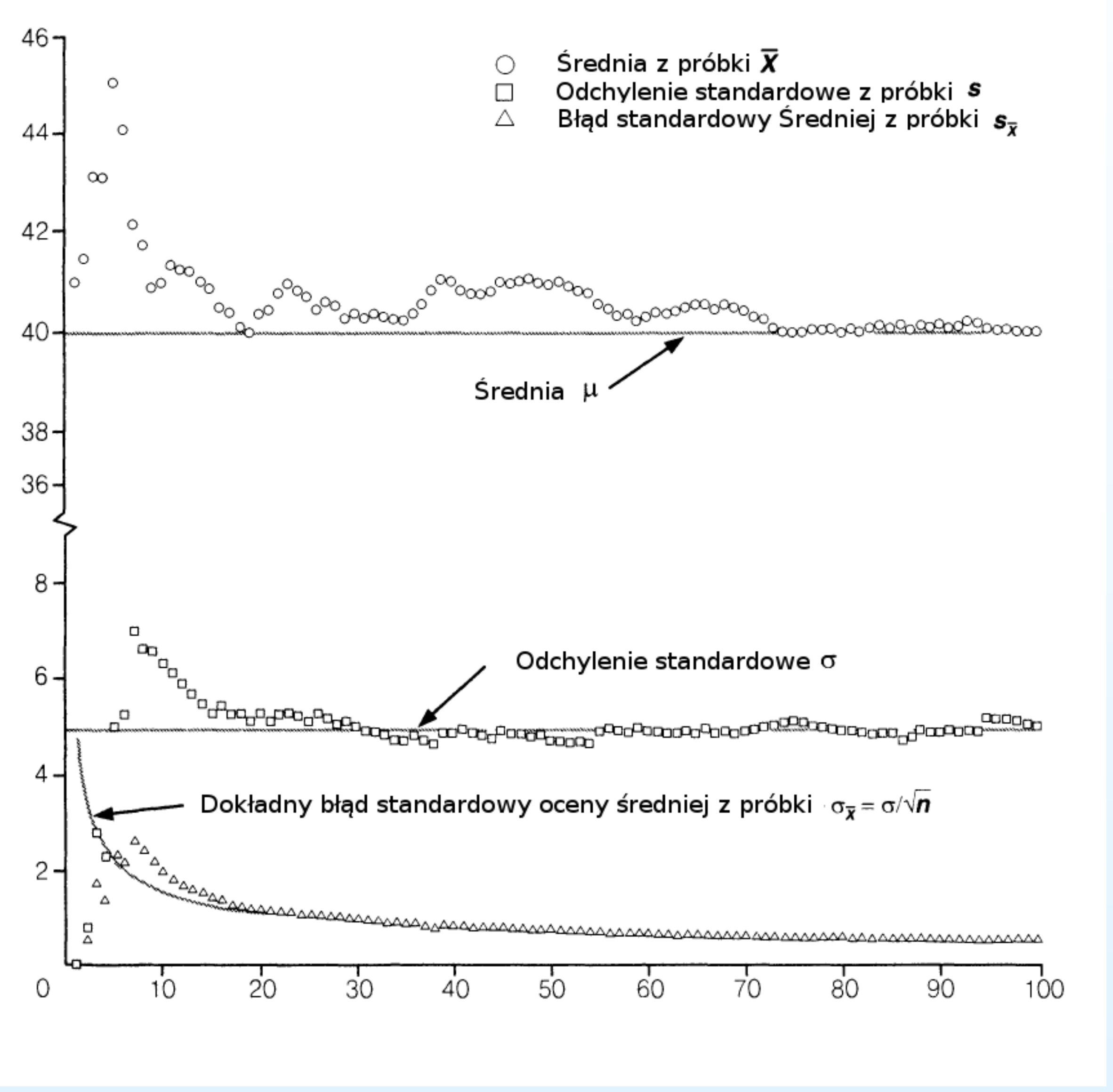
$$s_{\bar{X}} = \frac{s}{\sqrt{N}} = \sqrt{\frac{\sum(\bar{X}_i - \bar{X})^2}{N(N-1)}}$$

- Średnia w 95% przypdaków leży w przedziale $\bar{X}_i \pm 2s_{\bar{X}}$

Centralne twierdzenie graniczne

- Rozkład średnich z próbki jest zbliżony do rozkładu normalnego
- Średnia rozkładu średnich z próbki zgadza się ze średnią rozkładu początkowego
- Odchylenie standardowe rozkładu średnich z próbki zależy zarówno od odchylenia standardowego początkowego rozkładu jak i od wielkości próbki

Ilustracja


Średnia z próbki $\overline{X}$
Odchylenie standardowe z próbki s
Błąd standardowy Średniej z próbki $s_{\overline{X}}$
Średnia μ
Odchylenie standardowe σ
Dokładny błąd standardowy oceny średniej z próbki $\sigma_{\overline{X}} = \sigma/\sqrt{n}$
46
44
42
40
38
36
8
6
4
2
0
10
20
30
40
50
60
70
80
90
100

Przykład

- W artykule napisane, że w wyniku badania 20 osób średni rzut serca wyniosł $5{,}0\ l/min$ z odchyleniem standardowym $1{,}0\ l/min$
- 95 % zmiennej o rozkładzie normalnym mieści się w przedzile średnia $\pm 2s$
- Czy
  - prawie wszyscy pacjenci mieli rzut serca od 3 do $7\ l/min$?
  - średnia wartość rzutu serca (z prawdopodobieństwem 95 %) znajduje się z przedziale od $4{,}56$ do $5{,}44\ l/min$?

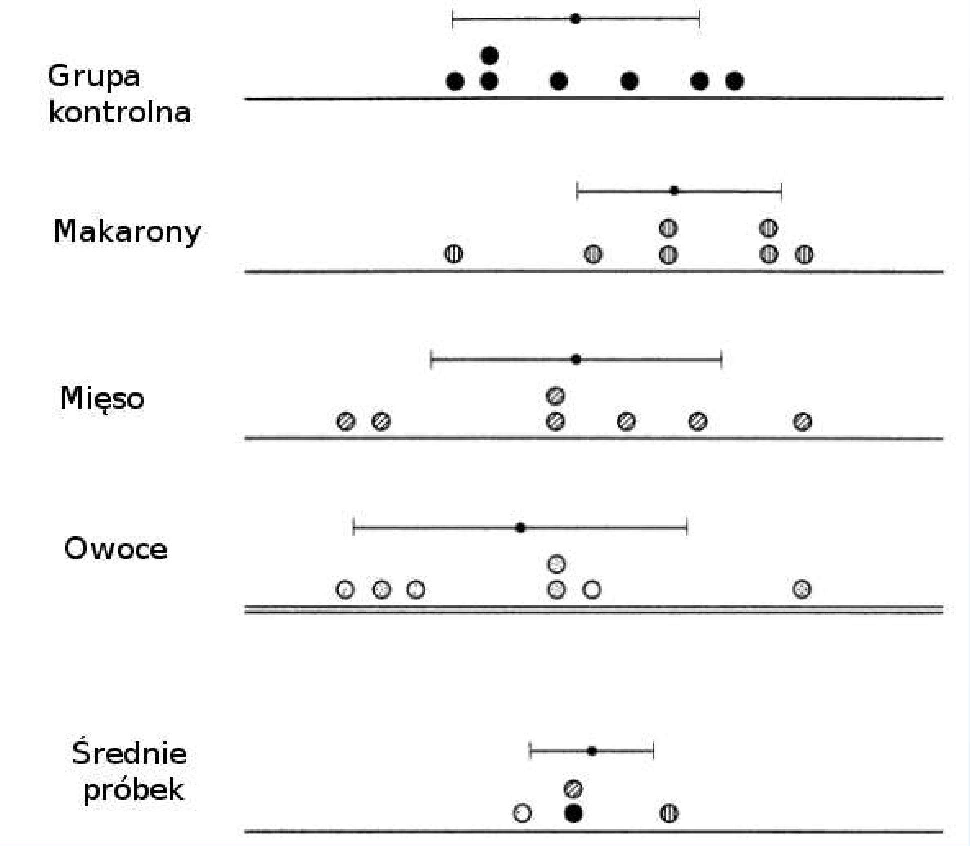
Poziom istotności różnicy

- Mamy kilka grup badanych jednostek
- Pytanie: czy różnice tych grup są statystycznie istonte?
  - hipoteza zerowa: różnice są przypadkowe
  - jakie jest prawdopodobieństwo otrzymać takie wyniki losowo?
    - Duze: hipoteza zerowa jest przyjmowana
    - Małe: hipoteza zerowa jest odrzucana
- Metoda: analiza wariancji

Losowe próbki z rozkładu normalnego

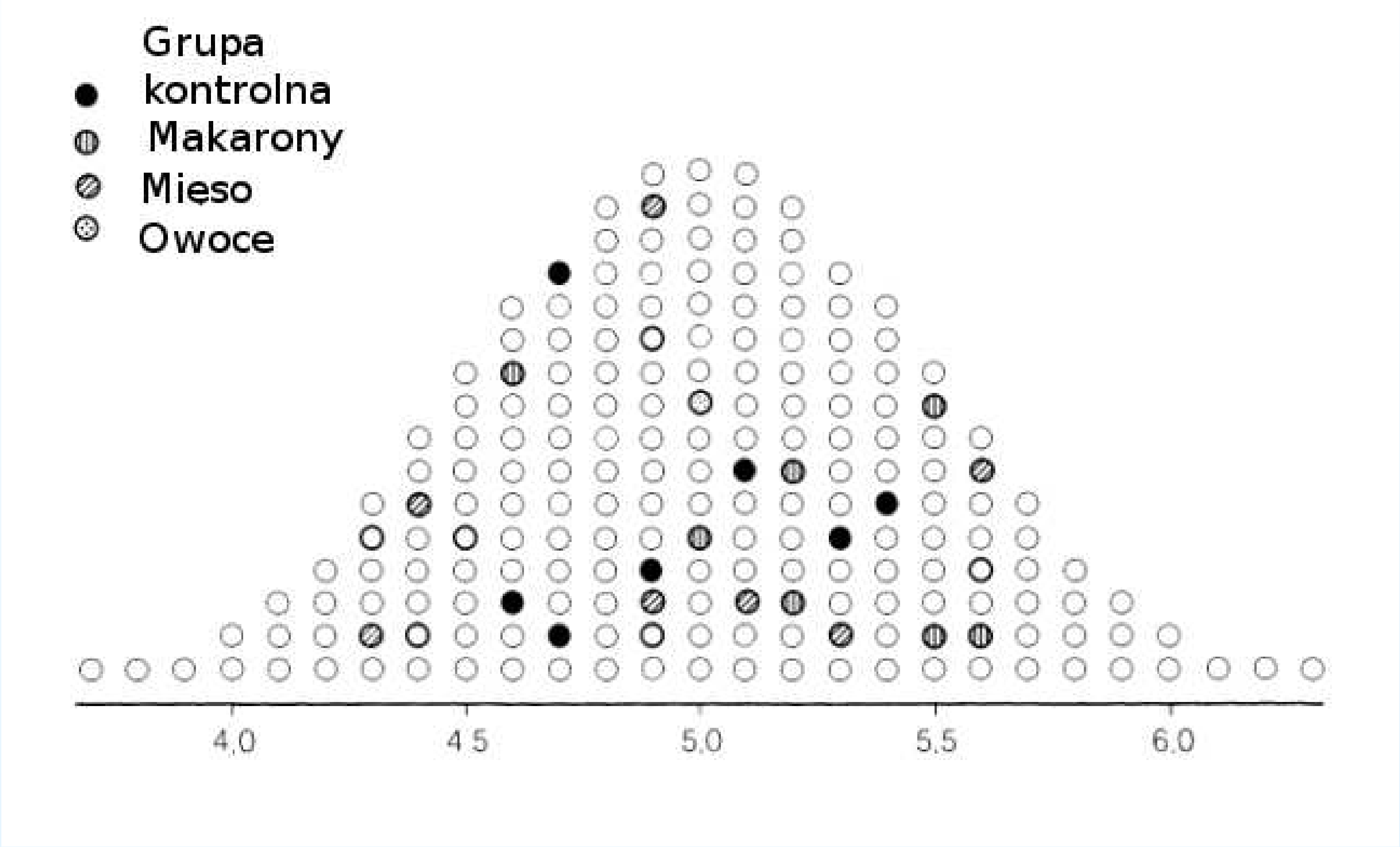
- W miasteczku (200 mieszkańców) badano wpływ diety na rzut serca
  - Wybrano losowo 28 mieszkańców
  - Podzielono na 4 grupy po 7 osób
    - Pierwsza grupa nie zmieniała diety
    - Druga jadła same makarony
    - Trzecia — mięso
    - Czwarta — owoce
  - po miesiącu zmierzono rzut serca

Wyniki



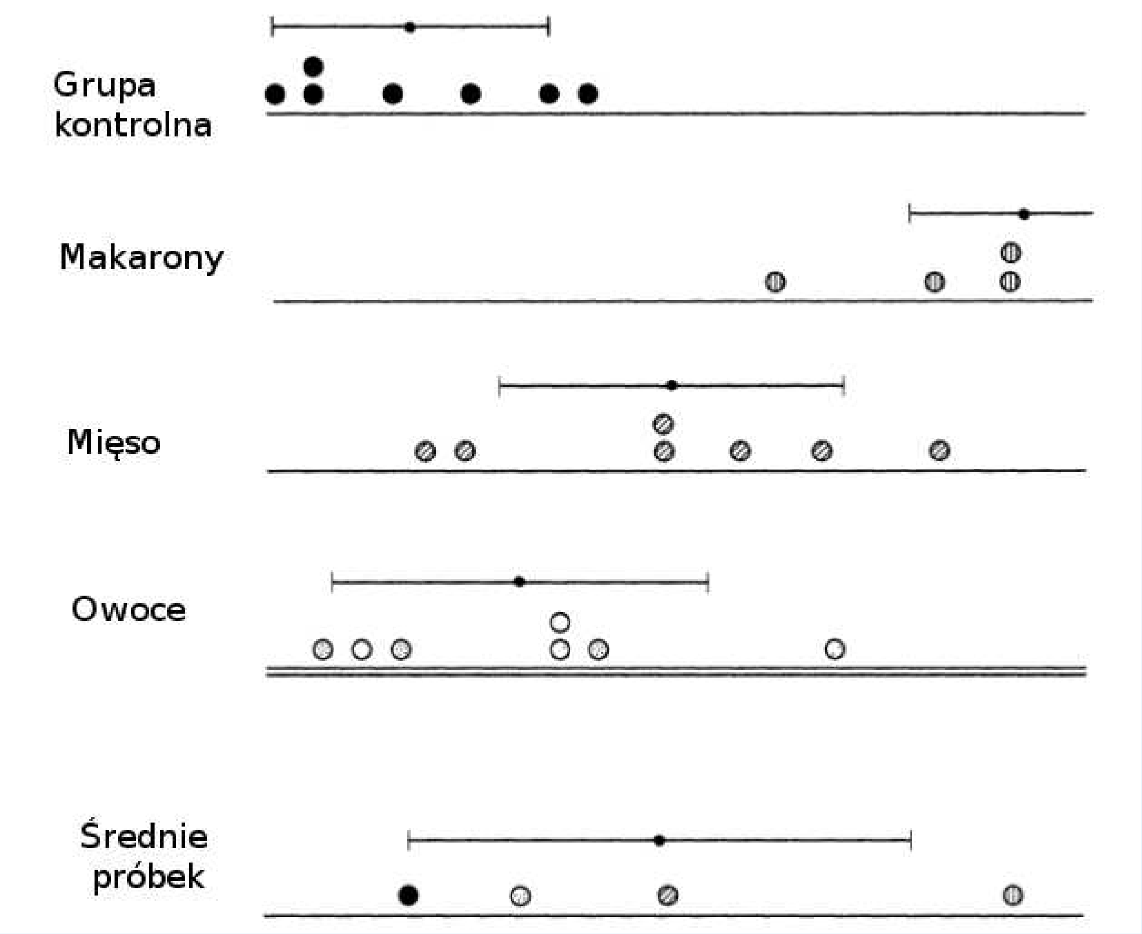
- Czy dieta ma wpływ na rzut serca?
- Hipoteza zerowa: nie ma wpływu.
- Jakie jest prawdopodobieństwo wylosować takie wyniki?

Całe miasteczko



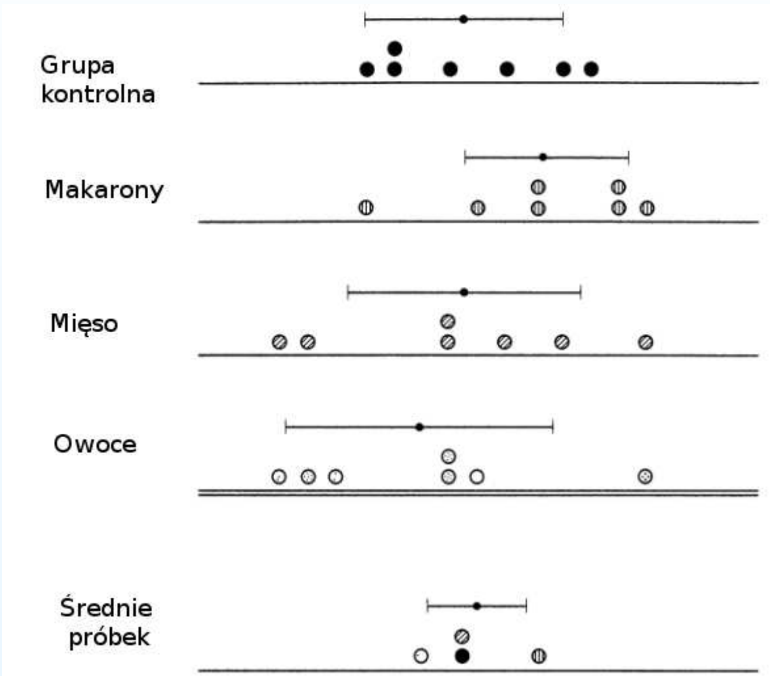
- Nasze grupy są losowymi próbkami z całości

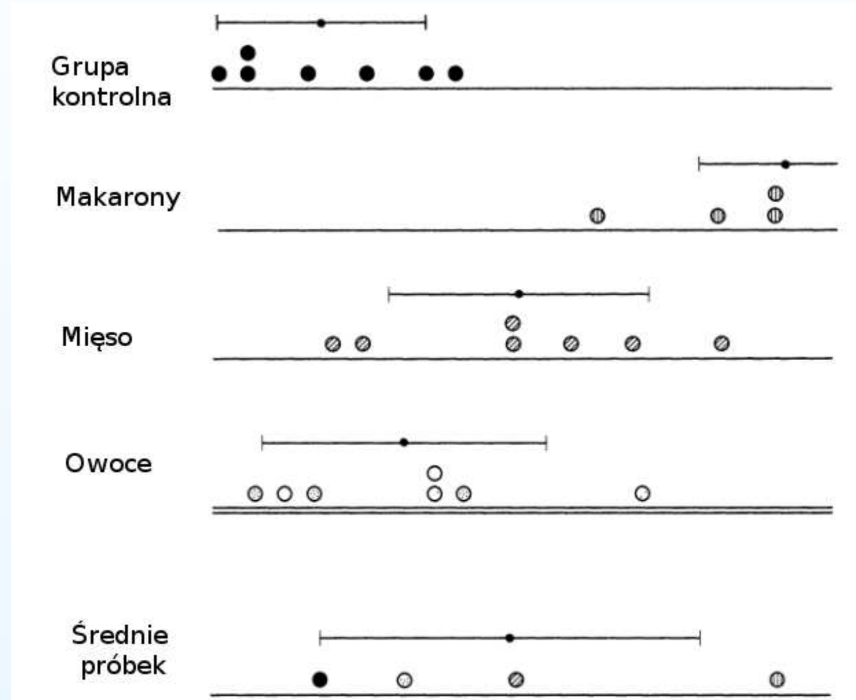
Czy takie wyniki się różnią?



- Chyba się różnią?

Porównywanie





- Czemu uważamy, że różnice w drugim przypadku są istotne?

Analiza wariancji

- Porównujemy wariancje
- Założenie: rozkład normalny
- Kriterium parametryczne

Dwa sposoby na obliczenie oceny wariancji

- Na podstawie wariancji grup (wariancja wewnętrzna)

$$s^2_{\mathrm{wewn}} = \frac{1}{4}\left(s^2_{\mathrm{kontr}} + s^2_{\mathrm{makar}} + s^2_{\text{mięso}} + s^2_{\mathrm{owoce}}\right)$$

- Na podstawie wariancji oceny średnich (wariancja międzygrupowa)

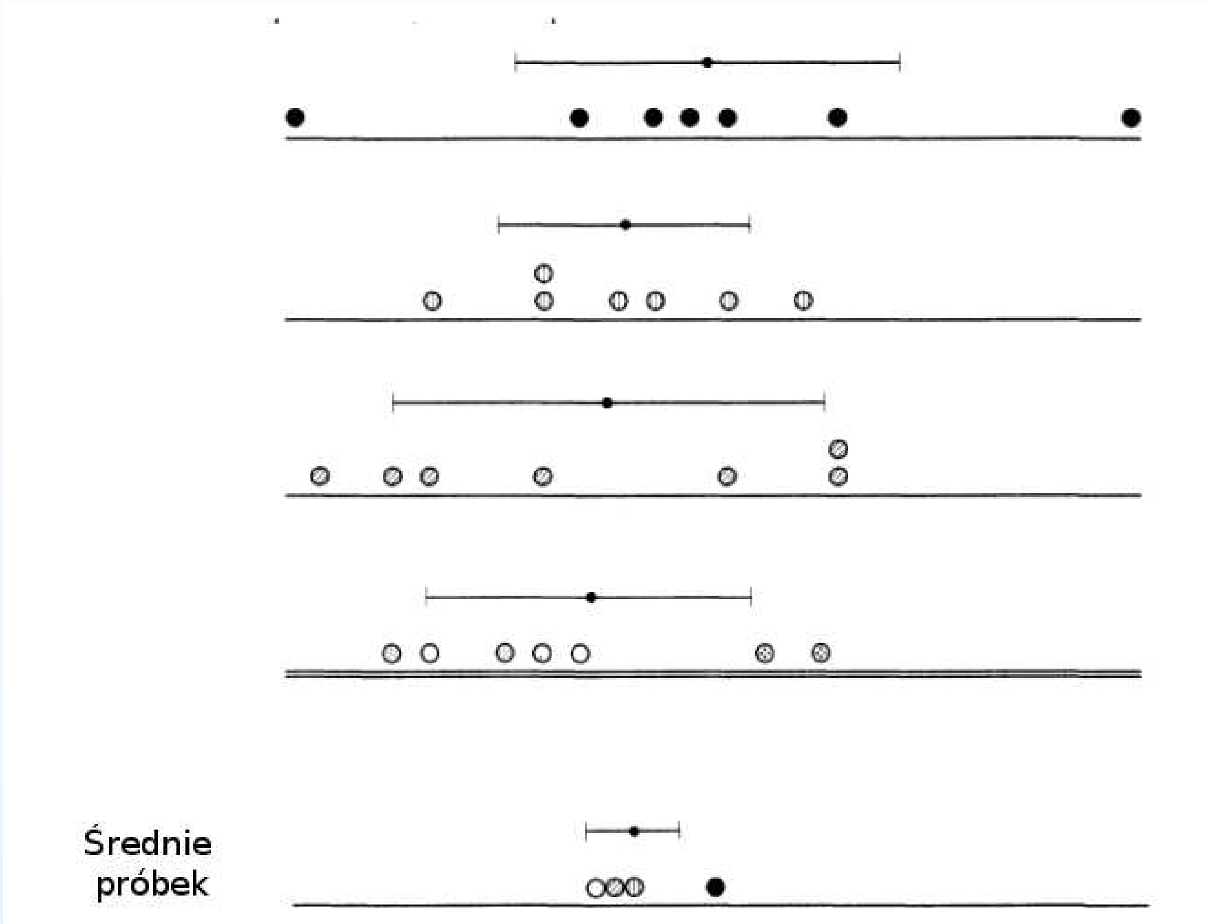
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} \Rightarrow s_{\mathrm{mg}} = N s^2_{\bar{X}}$$

- Jeżeli grupy są losowymi podzbiorami rozkładu normalnego, to obydwie oceny dają przybliżono tez sam wynik.

Test f (test Fishera)

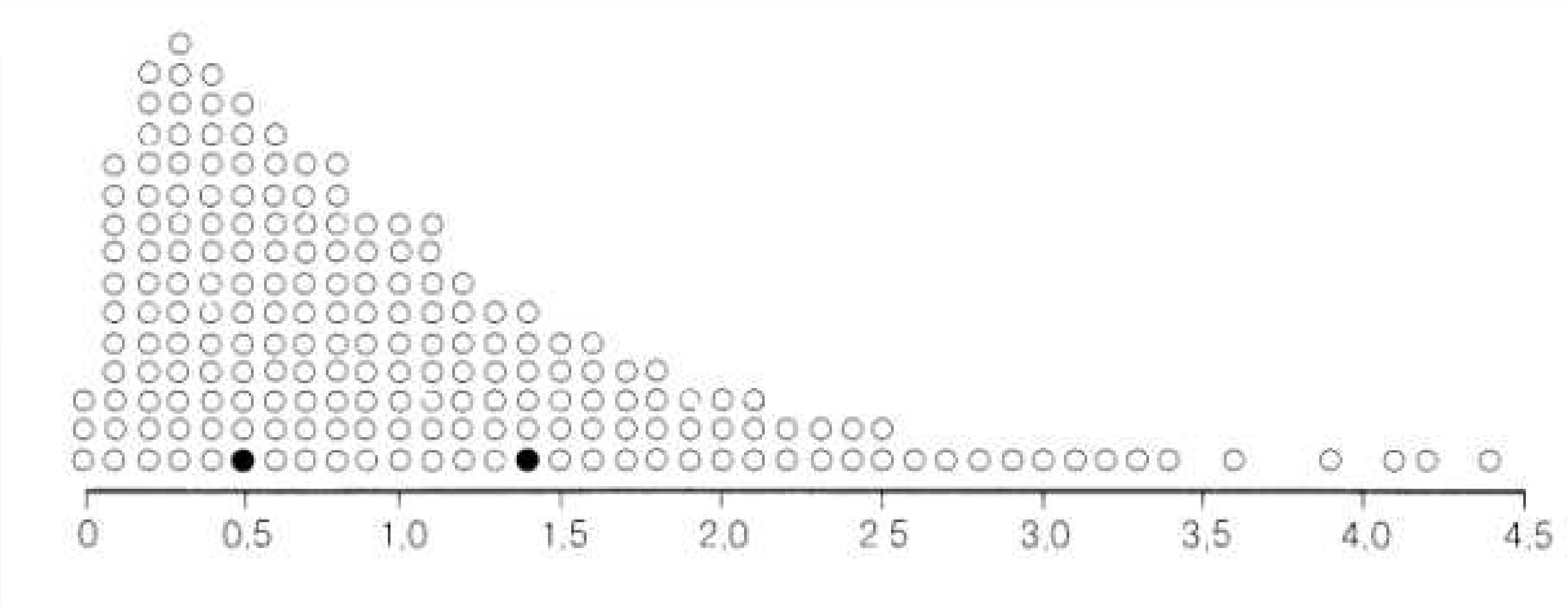
- $F = \dfrac{s^2_{\mathrm{mg}}}{s^2_{\mathrm{wewn}}}$
  - jeżeli $F \approx 1$, to hipotezę zerową należy przyjąć
  - jeżeli $F \gg 1$ (znacznie więcej), to hipotezę zerową należy odrzucić
    - Dla danych eksperymentu F jest blisko do jedynki
    - Dla danych „poprawionych” $F = 68,0$.

Wartość krytyczna F



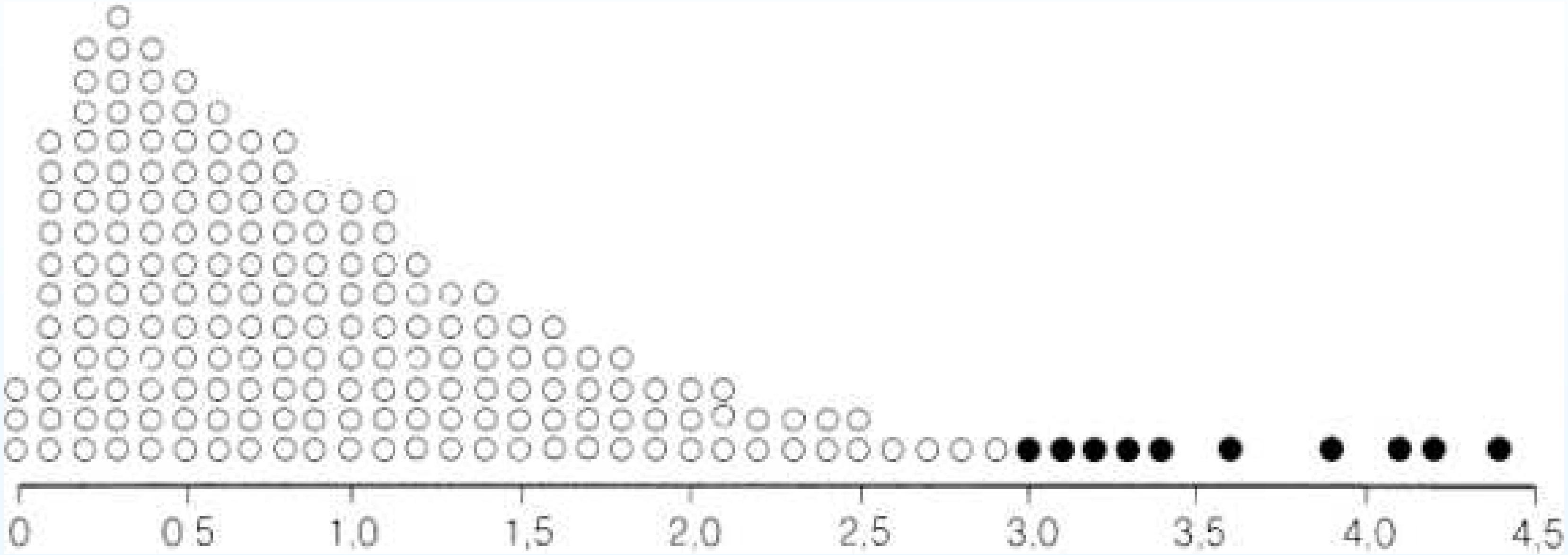
- Wartości F dla losowych próbek będą się różnić
- Od jakiej wartości F należy uważać za dużą?

Obliczymy F dla 200 próbek



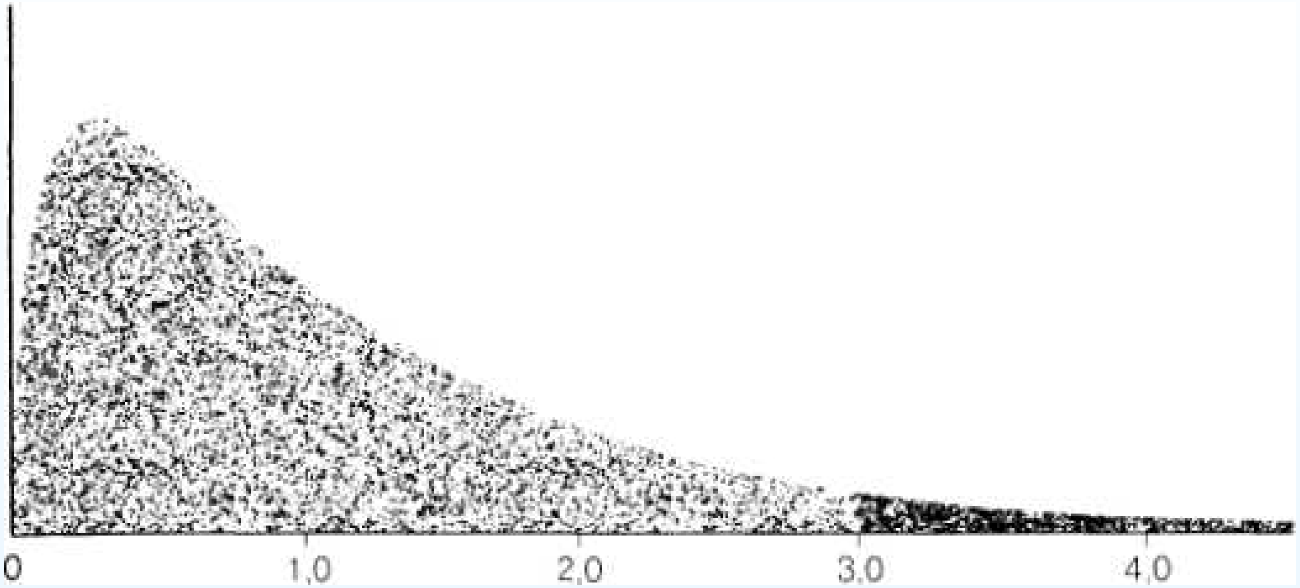
- Dla większości próbek F jest około jedynki (od 0 do 2)

Obliczymy F dla 200 próbek



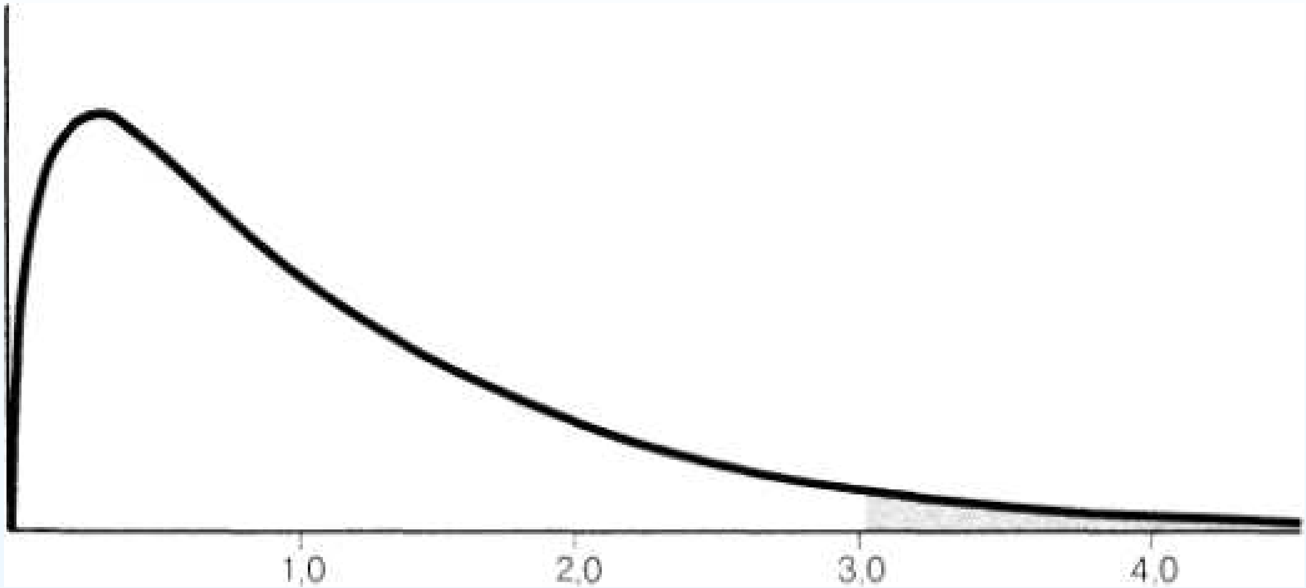
- Dla dziesięciu próbek $F \geqslant 3$
- Jeżeli będziemy odrzucać hipotezę zerową dla $F \geqslant 3$, to pomylimy się tylko w 5% przypadków
- $F = 3$ nazywa się *wartością krytyczną*
- $\alpha = 0{,}05$ (5%) nazywa się *poziomem ufności*

Obliczymy F dla wszystkich 7-osobowych próbek



- Wartość krytyczna wynosi około $3{,}01$ na poziomie ufności 5%

Gdyby liczebość całej populacji dążyła do ∞



- Wartość krytyczna wynosi $3{,}01$ na poziomie ufności 5%

Ogólny przypadek

- Wartość krytyczna wynosi zależy od
  - poziomu ufności (5%, 1%)
  - liczby stopni swobody $\nu_1 = m - 1$, $\nu_2 = m(n - 1)$, gdzie m jest ilością grup, n jest liczebnością każdej grupy
- Wartość krytyczna może być znaleziona z tablic $F_\alpha(\nu_1, \nu_2)$.
- Przykład tablicy:
  NIST/SEMATECH e-Handbook of Statistical Methods
- W naszym przykładzie 4 grupy po 7 osób

Założenia

- Założenia modelu matematycznego:
  - każda próbka jest niezależna od pozostałych
  - każda próbka wylosowana z całej populacji statystycznej
  - populacja statystyczna ma rozkład Gaussa
  - wariancje wszystkich próbek są równe
- Jeżeli jedno z założeń jest mocno naruszono, to nie można korzystać z metody

Przykład pierwszy

- Czy poprawne leczenie pozwala na skrócenie czasu hospitalizacji?
- D. E. Knapp, D. A. Knapp, M. K. Speedie, D. M. Yaeger, C. L. Baker: Relationship of inappropriate drug prescribing to increased length of hospital stay, *Am. J. Hosp. Pharm.*, **36**:1334–1337, 1979.

Organizacja badań

- Zbadano historie choroby pacjentów szpitala z ostrym odmiedniczkowym zapaleniem nerek
    - Badanie obserwacyjne (nie eksperymentalne)
    - Badanie retrospektywne (nie prospektywne)

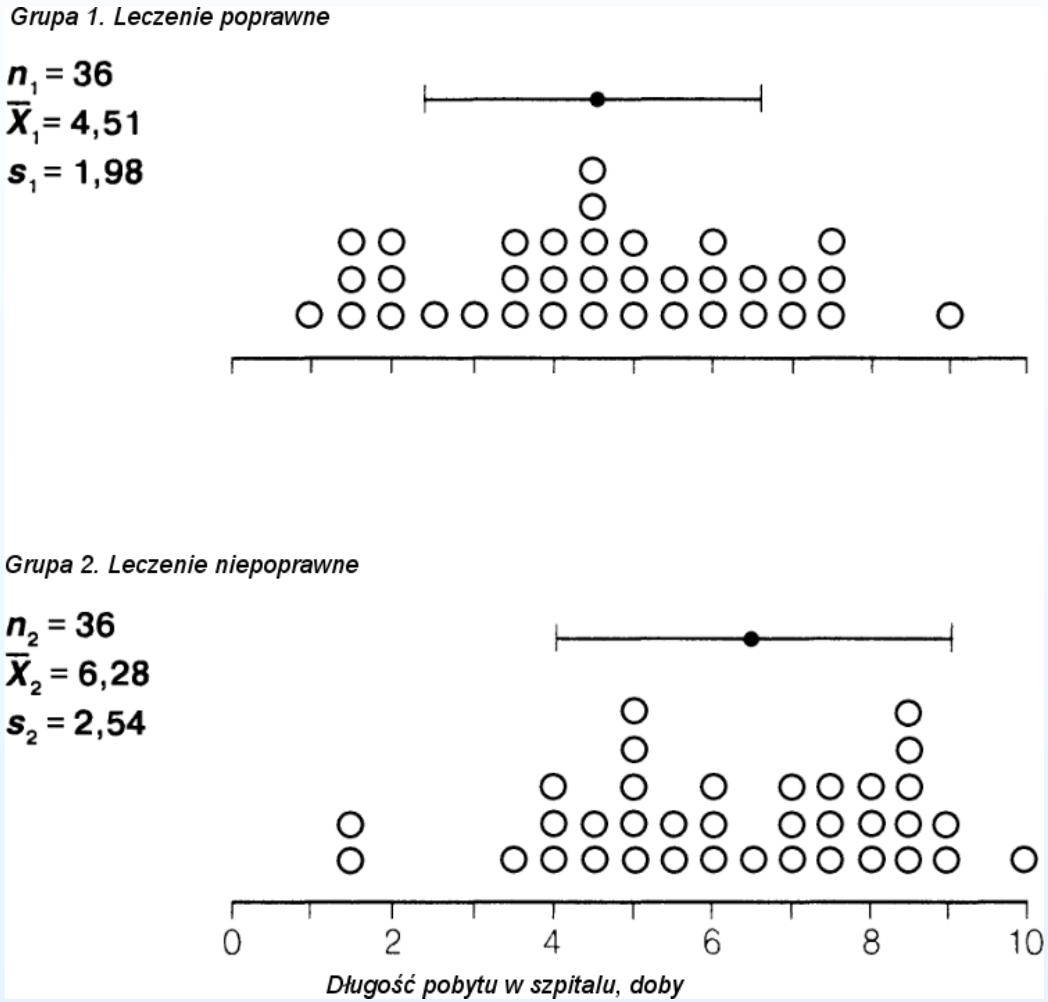
Kriteria włączenia pacjenta do badań

1. Diagnoza przy wypisie ze szpitala: *ostre odmiedniczkowe zapalenie nerek*
2. Przy przyjęciu: bole w dolnej części pleców, gorączka powyżej $37{,}8^\circ$
3. Bakteriuria co najmniej 10000 żywych bakterii w 1 ml moczu, stwierdzona wrażliwość na antybiotyk
4. Wiek od 18 do 44 lat
5. Brak niewydolności nerek lub wątroby, a także chorób, wymagających leczenia chirurgicznego
6. Wypis ze szpitala w związku z poprawą stanu zdrowia

Kriterium poprawności leczenia

- Leczenie poprawne, jeżeli odpowiada zaleceniom "Physicians' Desk Reference"
  - Według tego kriterium pacjenci zostali podzieleni na dwie grupy:
    - Leczone poprawnie
    - Leczone niepoprawnie

Wyniki



- Obliczyć F.

Wnioski

- $F = 10{,}81$
- Wartość krytyczna na poziomie 1% $7{,}01$
- Czy leczeni poprawnie byli w szpitalu dzięki poprawnemu leczeniu? Niewiadomo. Czemu?

Przykład drugi

- Halotan czy Morfina podczas operacji na otwartym sercu
  - Halotan powoduje spadek ciśnienia tętniczego. Czy warto zastąpić morfiną?
- T. J. CONAHAN III, A. J. OMINSKY, H. WOLLMAN, R. A. STROTH: A prospectiverandom comparison of halothane and morphine for open-heart anesthe-sia: one year experience. *Anesthesiology,* **38**:528–535, 1973.

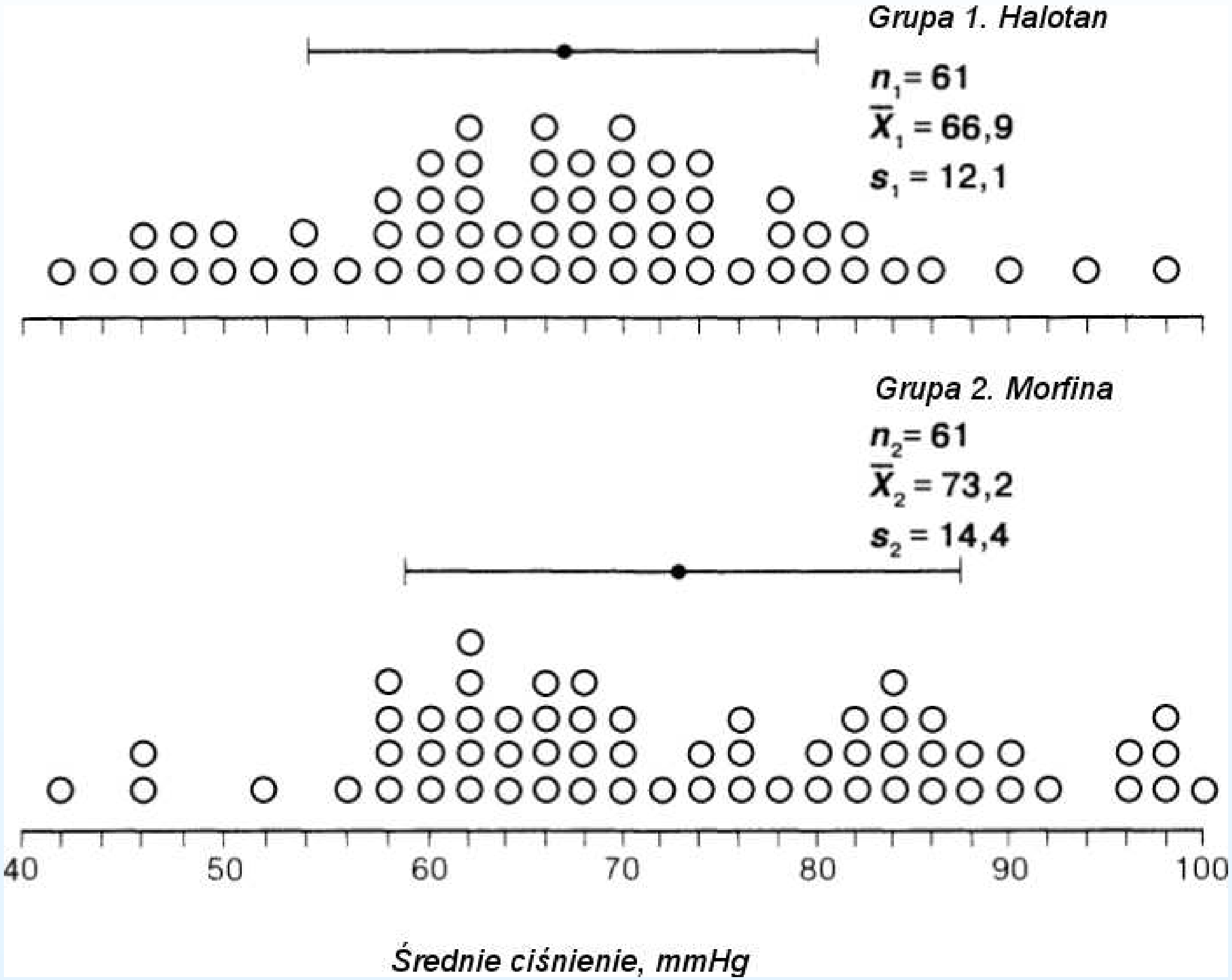
Organizacja badań

- Pacjenci, którzy nie mają przeciwwskazań wobec halotanu, ani wobec morfiny
- Sposób anestezji wybierany losowo
  - Badanie kontrolowane, prospektywne
  - Badanie eksperymentalne
- Ciśnienie tętnicze między początkiem anestezji a początkiem operacji mierzono wielokrotnie

$$C = \frac{C_{\text{skurcz}} - C_{\text{rozkurcz}}}{3} + C_{\text{rozskurcz}}$$

- W badaniu wykorzystana najmniejsza wartość

Wyniki



- Obliczyć F.

Wnioski

- $F = 6{,}81$
- Wartość krytyczna na poziomie 5% $3{,}92$

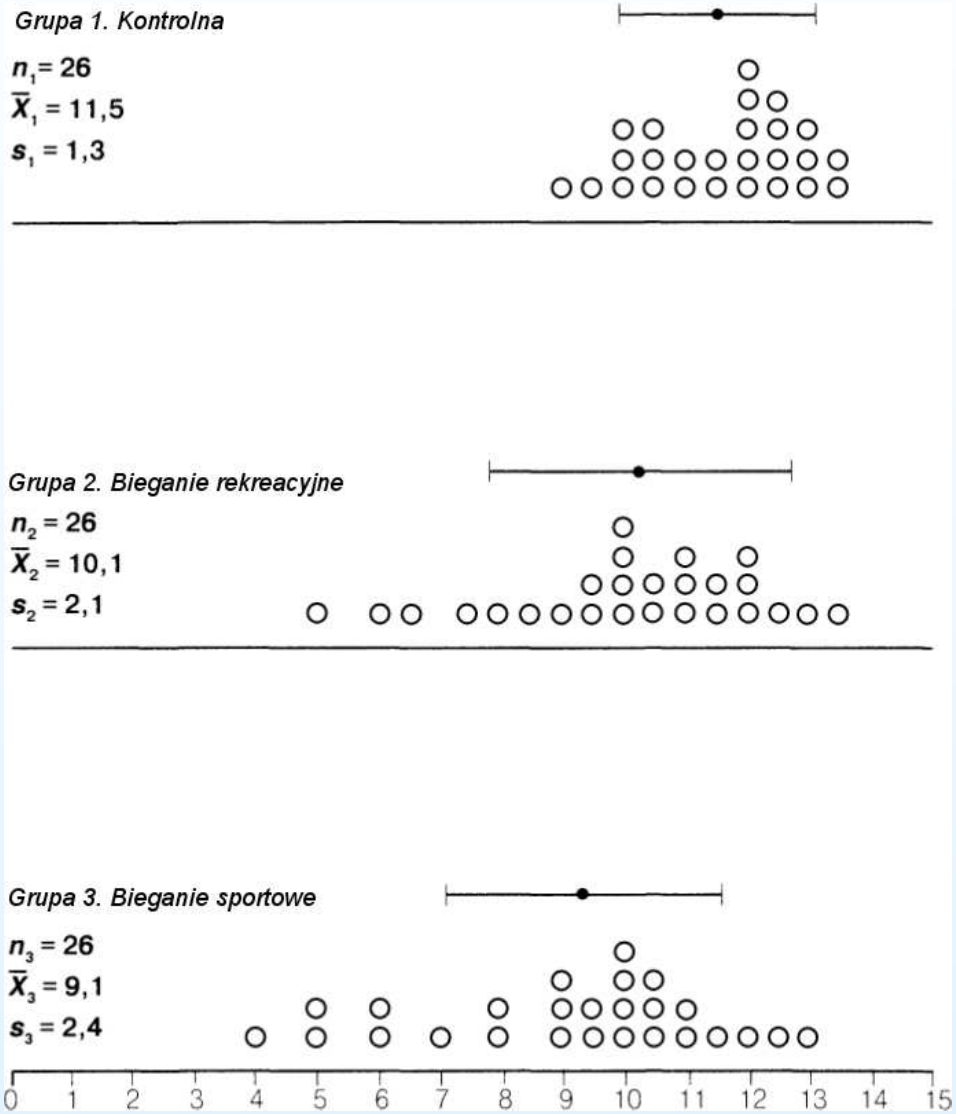
Przykład trzeci

- Sport a zaburzenia miesiączkowania
  - Czy uprawianie biegania ma związek z zaburzeniem miesiączkowania?
- E. DALE, D. H. GERLACH, A. L. WILHITE.: Menstrual dysfunction in distance runners. *Obs. GynecoL,* **54**:47–53, 1979.

Organizacja badań

- 78 młodych kobiet, podzielonych na trzy grupy po 26 osób
  1. Grupa kontrolna: kobiety, nie uprawiające sportu
  2. Kobiety, uprawiające bieganie rekreacyjnie (jogging), przebiegają od 6 do 48 kilometrów tygodniowo
  3. Kobiety, uprawiające bieganie sportowo, przebiegają ponad 48 kilometrów tygodniowo
- Obliczona roczna ilość menstruacji

Wyniki



Grupa 1. Kontrolna
$n_1 = 26$
$\bar{X}_1 = 11,5$
$s_1 = 1,3$
Grupa 2. Bieganie rekreacyjne
$n_2 = 26$
$\bar{X}_2 = 10,1$
$s_2 = 2,1$
Grupa 3. Bieganie sportowe
$n_3 = 26$
$\bar{X}_3 = 9,1$
$s_3 = 2,4$
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Wnioski

- $F = 9{,}48$
- Wartość krytyczna na poziomie 1% $4{,}90$