Statystyka Opisowa z Demografią oraz Biostatystyka

Test t Studenta

Aleksander Denisiuk

denisjuk@euh-e.edu.pl

Elbląska Uczelnia Humanistyczno-Ekonomiczna

ul. Lotnicza 2

82-300 Elbląg

Test t Studenta

Najnowsza wersja tego dokumentu dostępna jest pod adresem

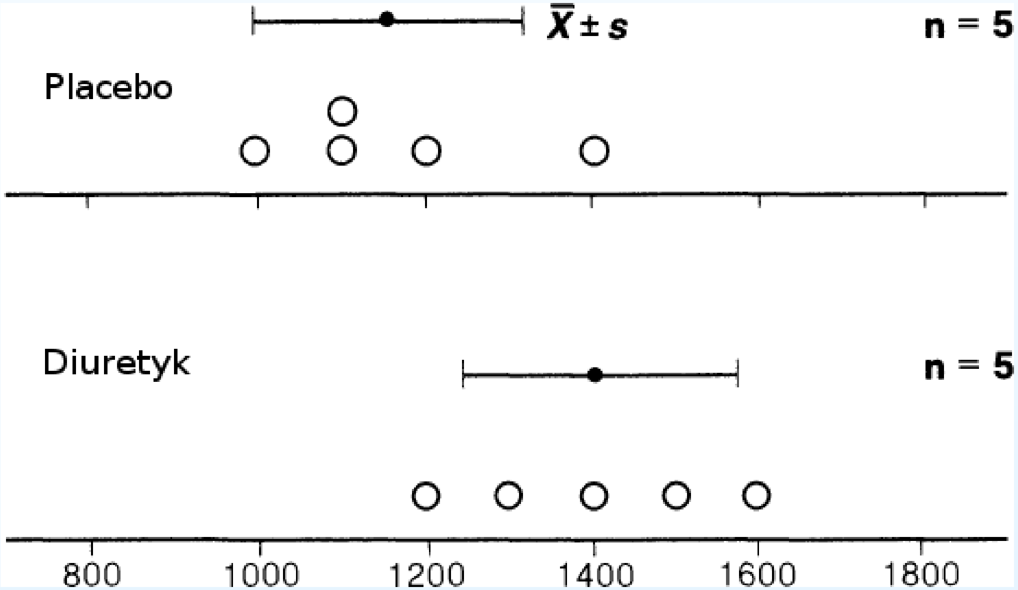
http://denisjuk.euh-e.edu.pl/

Test t-Studenta

- Służy do porównania *dwóch* grup
- Jest bardzo popularny
- Często niepoprawnie stosuje się do porównania więcej niż dwóch grup

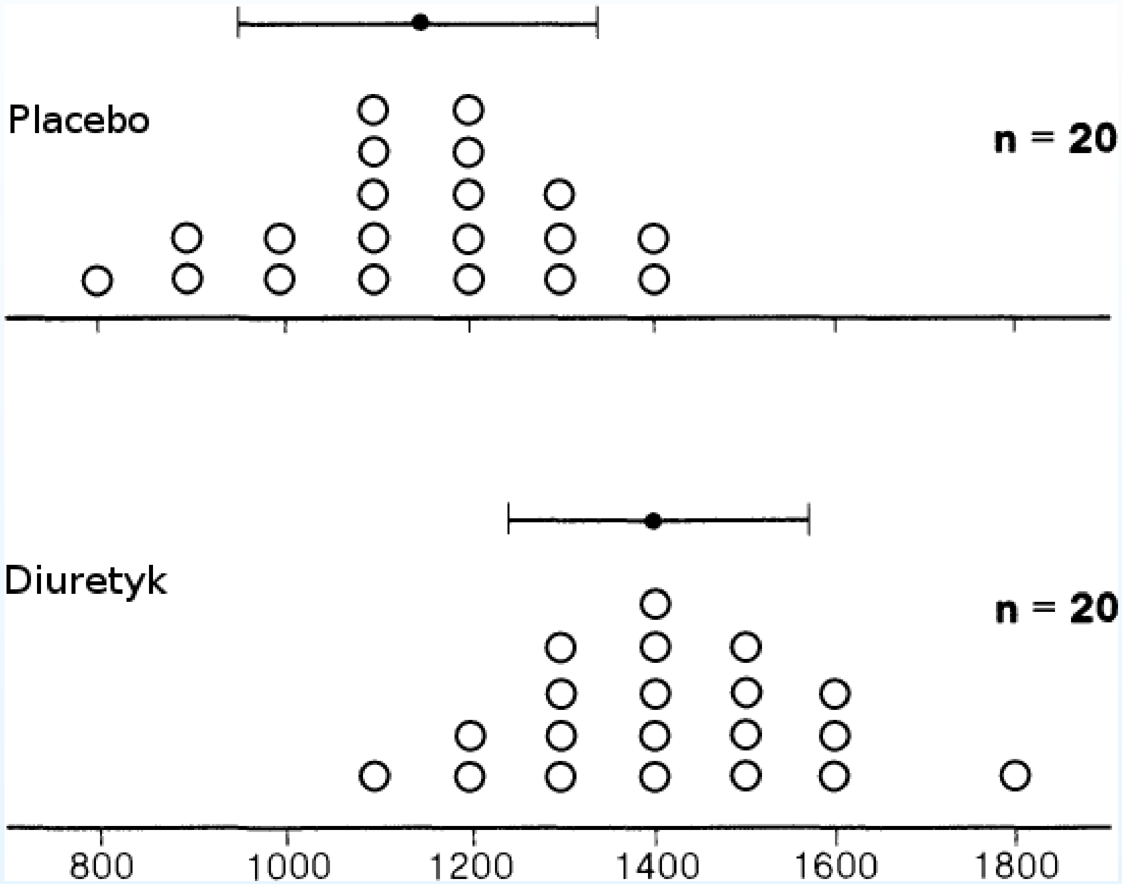
Zasada testu. Przykład

- Badamy skuteczność nowego diuretyka
- Dwie grupy po pięć osób
- Mieżono dobową diurezę



- Czy nowy diuretyk jest skuteczny?

Większe grupy



- Czemu mamy większe zaufanie do drugiego doświadczenia?

Bo. . .

- Dokładność oceny wartości średniej z próby zależy od liczebności próby:

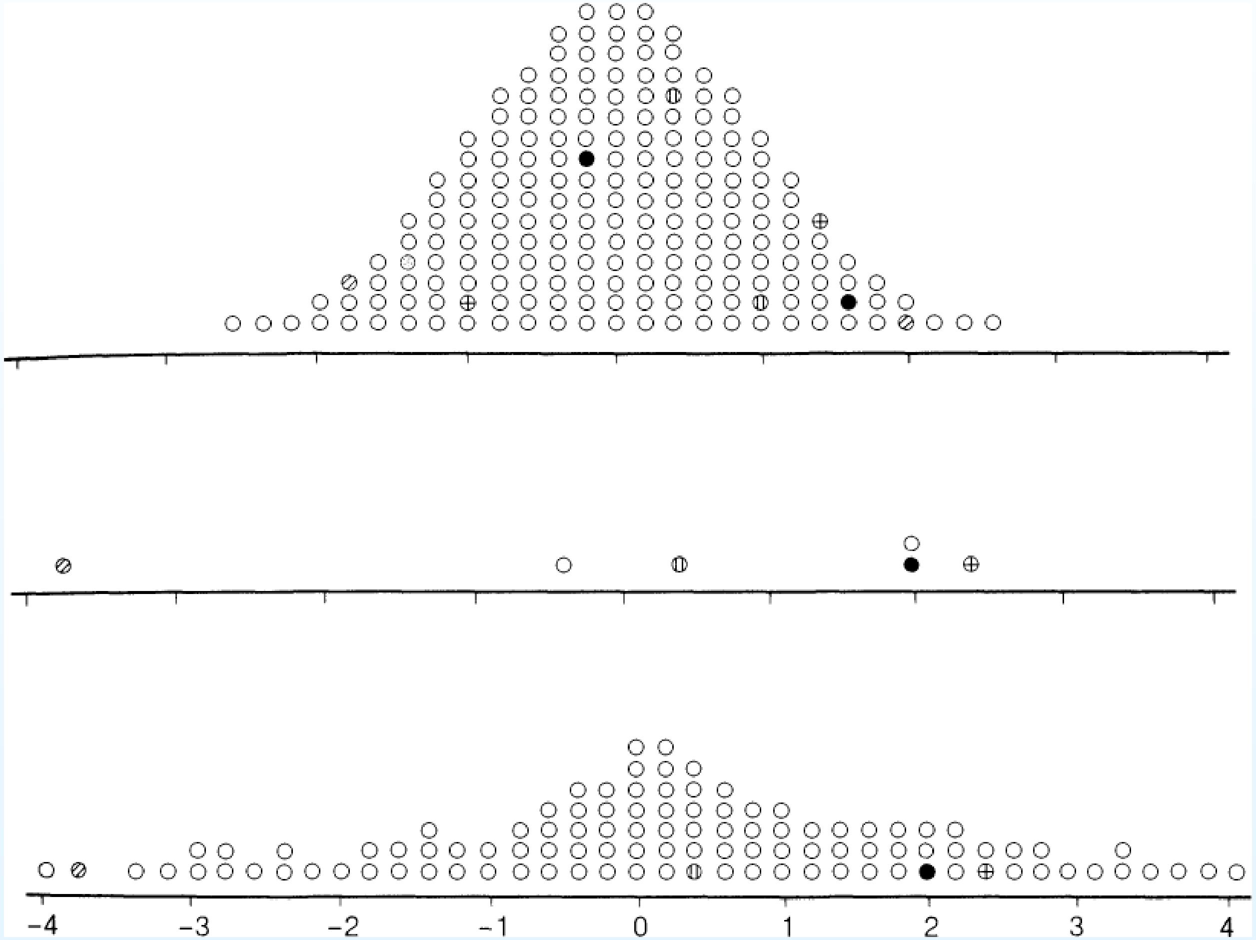
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

- Im większa jest próba, tym mniejszy jest błąd standardowy.
- Tym mniejsze jest prawdopodobieństwo hipotezy zerowej (że obydwie próbki pochodzą z tej samej *normalnej* zbiorowości statystycznej).

Test t

- $t = \dfrac{\text{różnica średnich z próbki}}{\text{błąd standardowy różnic średnich z próbki}}$
  - jeżeli $t \approx 0$, to hipotezę zerową należy przyjąć
  - jeżeli $|t| \gg 1$ (znacznie więcej), to hipotezę zerową należy odrzucić

Błąd standardowy różnic średnich z próbki



- $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$

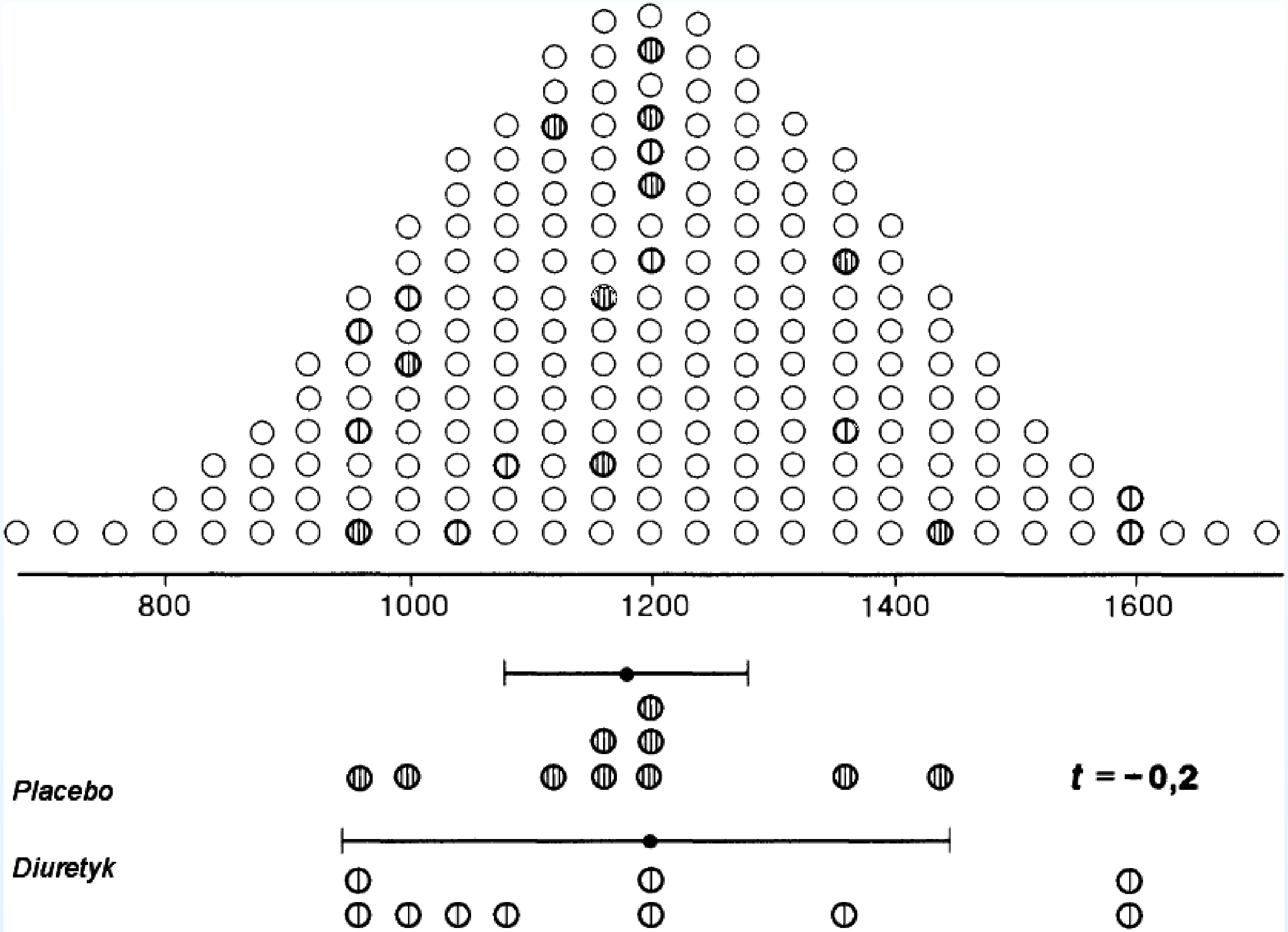
Ocena błędu standardowefo różnic średnich z próbki

- $\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$
- $s^2_{X-Y} = s^2_X + s^2_Y$
- $s^2_{\bar{X}-\bar{Y}} = s^2_{\bar{X}} + s^2_{\bar{Y}}$
- $s_{\bar{X}-\bar{Y}} = \sqrt{s^2_{\bar{X}} + s^2_{\bar{Y}}}$

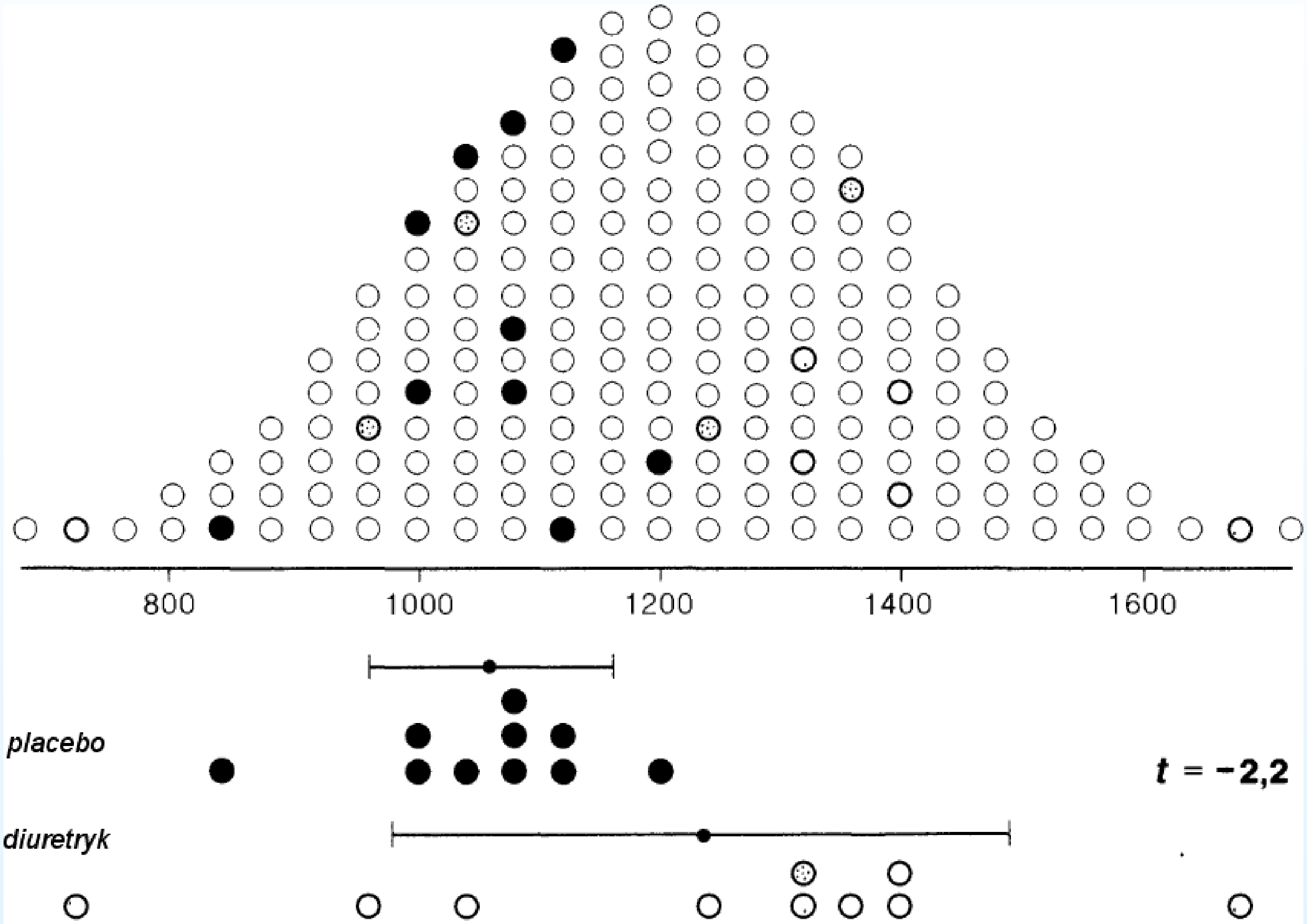
Ocena testu t

- $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2_{\bar{X}_1} + s^2_{\bar{X}_2}}}$
- $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}}$
- jeżeli obydwie próbki zostały pobrane z tej samej zbiorowości, to s_1^2 oraz s_2^2 są ocenami tej samej wariancji σ
- zastępujemy je przez ocenę zbiorczą: $s^2 = \frac{s_1^2 + s_2^2}{2}$ (ta sama wielkość próbek)
  - $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n} + \frac{s^2}{n}}}$

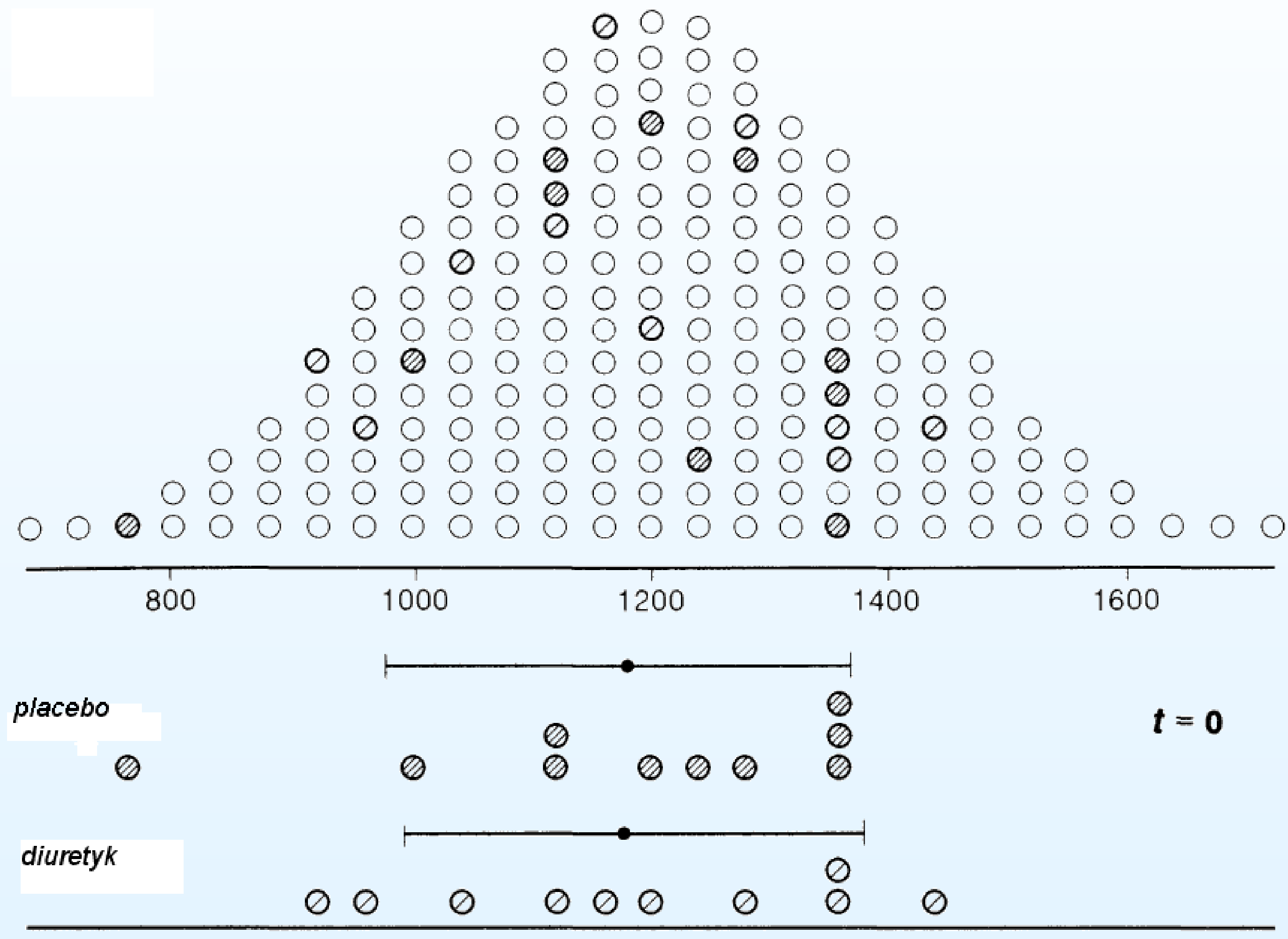
Wartość graniczna t



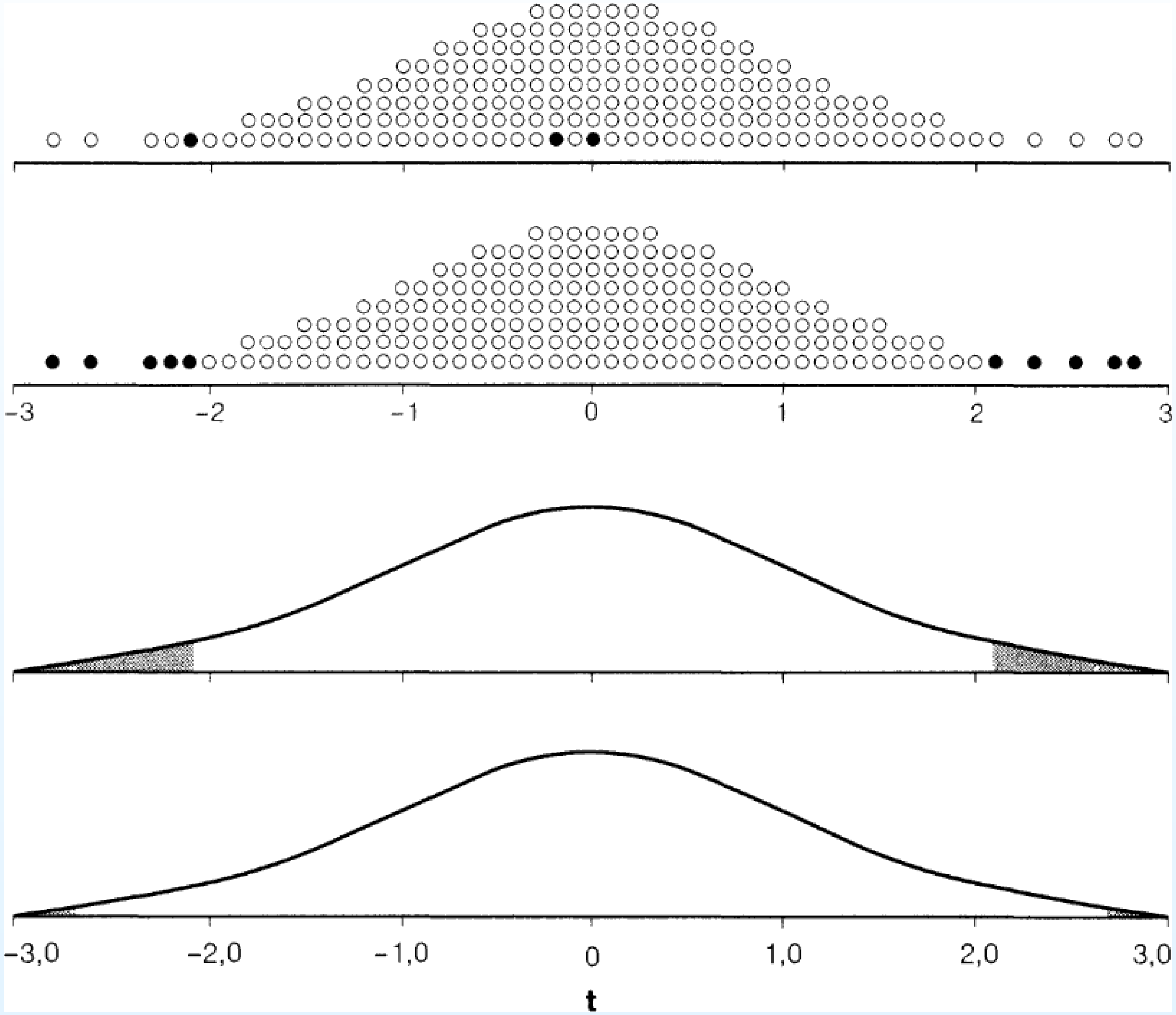
Inne dwie próbki



Jeszcze dwie probki



Dwieście i więcej prób



Wartość krytyczna t

- zależy od poziomu ufności α oraz od liczby stopni swobody
- dla dwóch próbek wielkości n

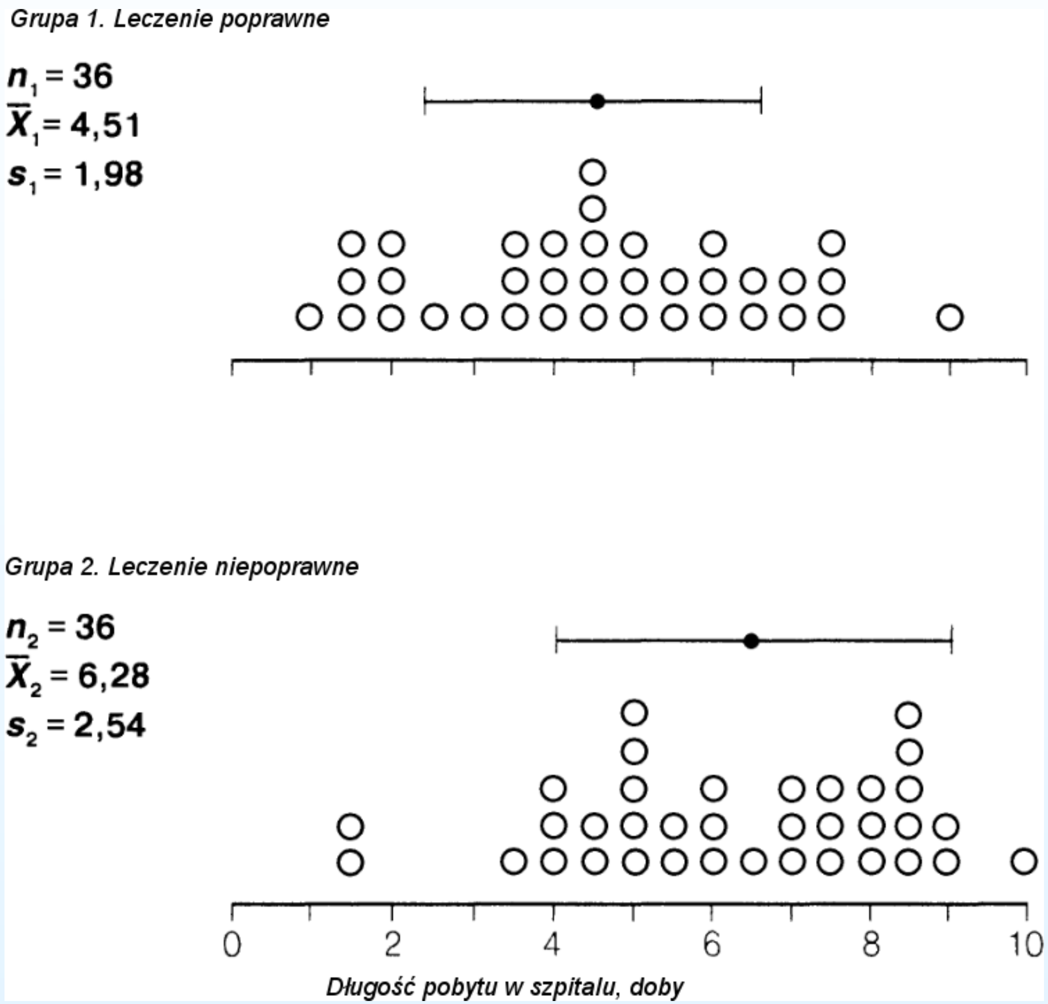
$$\nu = 2(n-1)$$

- test dwustronny (two-sided test)
- Przykład tablicy:
  NIST/SEMATECH e-Handbook of Statistical Methods

Próbki różnych wielkości

- ocena zbiorza wariancji $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
- $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$
- liczba stopni swobody $\nu = n_1 + n_2 - 2$

Przykład z zapaleniem nerek ponownie

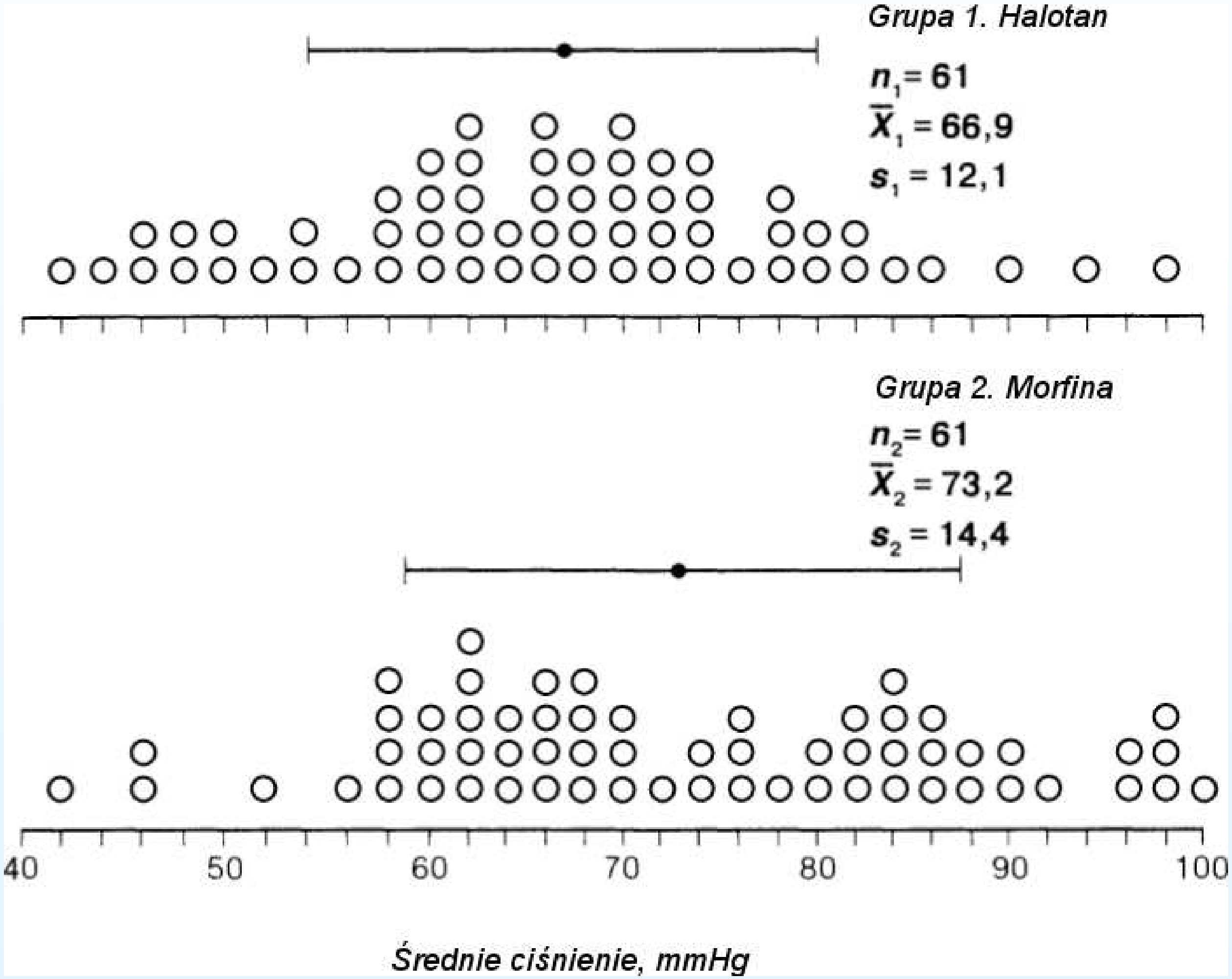


- Obliczyć t.

Wnioski

- $t = -3{,}30$
- Wartość krytyczna na poziomie 1% $6{,}648$
- Różnica długości terminu pobytu w szpitalu statystycznie znacząca.

Halotan czy Morfina?



Wnioski

- $t = -2{,}607$
- Wartość krytyczna na poziomie 5% $1{,}980$
- Morfina obniża ciśnienie krwi w mniejszym stopniu, niż halotan

Wskaźnik sercowy

- W badaniu został zmierzony również wskaźnik sercowy

Halotan ($n = 9$)		Morfina ($n = 16$)	
$\bar{X}_1$	s_1	$\bar{X}_2$	s_2
2,08	1,05	1,75	0,88

- $s^2 = 0{,}98$
- $t = 0{,}84$
- $\nu = 23$
- Wartość krytyczna na poziomie 5% $2{,}069$
- Brak różnic statystycznych

Popularny błąd w stosowaniu testu t

- Test t Studenta opracowany jest dla porównywania dwóch próbek.
- Błędem jest stosowanie go dla większej ilości próbek.
- Przykładowo:
  - Badano wpływ farmacutyków A i B na poziom glukozy w plasmie
  - Trzy grupy: A, B, oraz placebo
  - Trzy testy Studenta: dla każdej pary
  - W jednym z testów t jest wystarczająco duże
  - Stwierdza się, że $P < 5\%$
  - Błąd: prawdopodobieństwo blędnie odrzucić hipotezę zerową jest znacznie większe!

Wpływ wielokrotnych porównań

- Jeżeli prawdopodobieństwo pomyłki w porównywaniu każdej pary grup jest $P(= 0{,}05)$, to prawdopodobieństwo pomylić się w porównywaniu przynajmniej jednej z par wyniesie

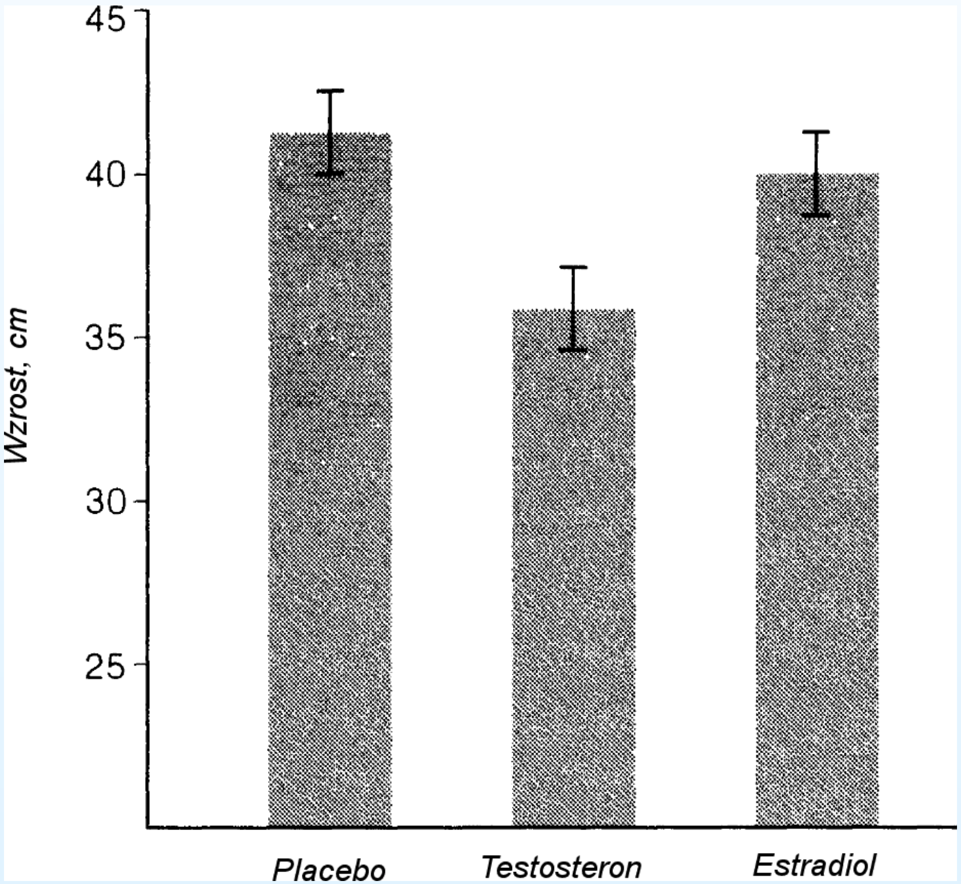
$$P' = 1 - (1 - P)^k,$$

gdzie k jest liczbą porównywań.

- W poprzednim przykadzie:
    - $P' = 1 - (1 - 0{,}05)^3 \approx 3 \cdot 0{,}05 = 15\%$

Przykład

- Badano wpływ ludzkich hormonów na wzrost Marsjan
  - my wiemy, że ludzkie hormoy nie działają na Marsjan
  - badacze nie wiedzą, otrzymali wyniki:



Test t dla przykładu

- grupy po 10 Marsjan
- test t wynosi:
  - placebo–testosteron $2{,}39$
  - placebo–estradiol $0{,}93$
  - testosteron–estradiol $1{,}34$
- $\nu = 18$
- wartośc krytyczna na poziomie 5%: $2{,}101$
- wnioski:
  1. testosteron powoduje zmniejszenie wzrostu
  2. estradiol nie powoduje zmiany wzrostu
  3. działanie testosteronu nie różni się od działania estradiolu
- Co jest nie tak?

Test F dla przykładu

- $F = 2{,}47$
- $\nu_1 = 2$, $\nu_2 = 27$
- wartośc krytyczna na poziomie 5%: $3{,}35$
- wniosek:
  - brak różnicy statystycznej

Podsumowanie

- Test Studenta wykorzystuje się tylko dla dwóch grup
- Jeżeli ilość grpu jest większa, należy korzystać z analizy wariancji
- Jeżeli test Studenta został wykorzystany dla większej ilości grup, to poziom ufności jest większy tyle razy, ile jest banadych grup

Test t-Studenta dla porównań wielokrotnych

- porównujemy więcej niż jedną grupę
- testu t nie można zastosować
- analiza wariancji odrzuca hipotezę zerową
- która z grup różni się statystycznie od innych?
  - metody porównań wielokrotnych
  - oparte na teście t-Studenta

Współczynnik korekcji Bonferroniego

- Carlo Emilio Bonferroni, 1892–1960
- Nierówność Bonferroniego
  - Jeżeli zastosować k razy test o poziomie ufności α, to pradwopodobieństwo błądenie odrzucić hipotezę zerową jest nie większe, niż $k\alpha$:

$$\alpha' < k\alpha$$

- a więc, żeby zapewnić poziom ufności α, dla każdego porównania trzeba przyjąć poziom α/k
  - w szczególności, tla trzech grup $0{,}05/3 \approx 1{,}7\%$

Modyfikacja korekcji Bonferroniego

- Dla dużej liczby porównań ($k \geqslant 8$) korekcja Bonferroniego staje się zbyt restrykcyjną
- Zmniejszmy restrykcyjność korekcji
  - $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$, gdzie s^2 jest oceną zbiorczą wariancji
  - jako oceny wariancji używamy oceny wariancji wewnętrznej

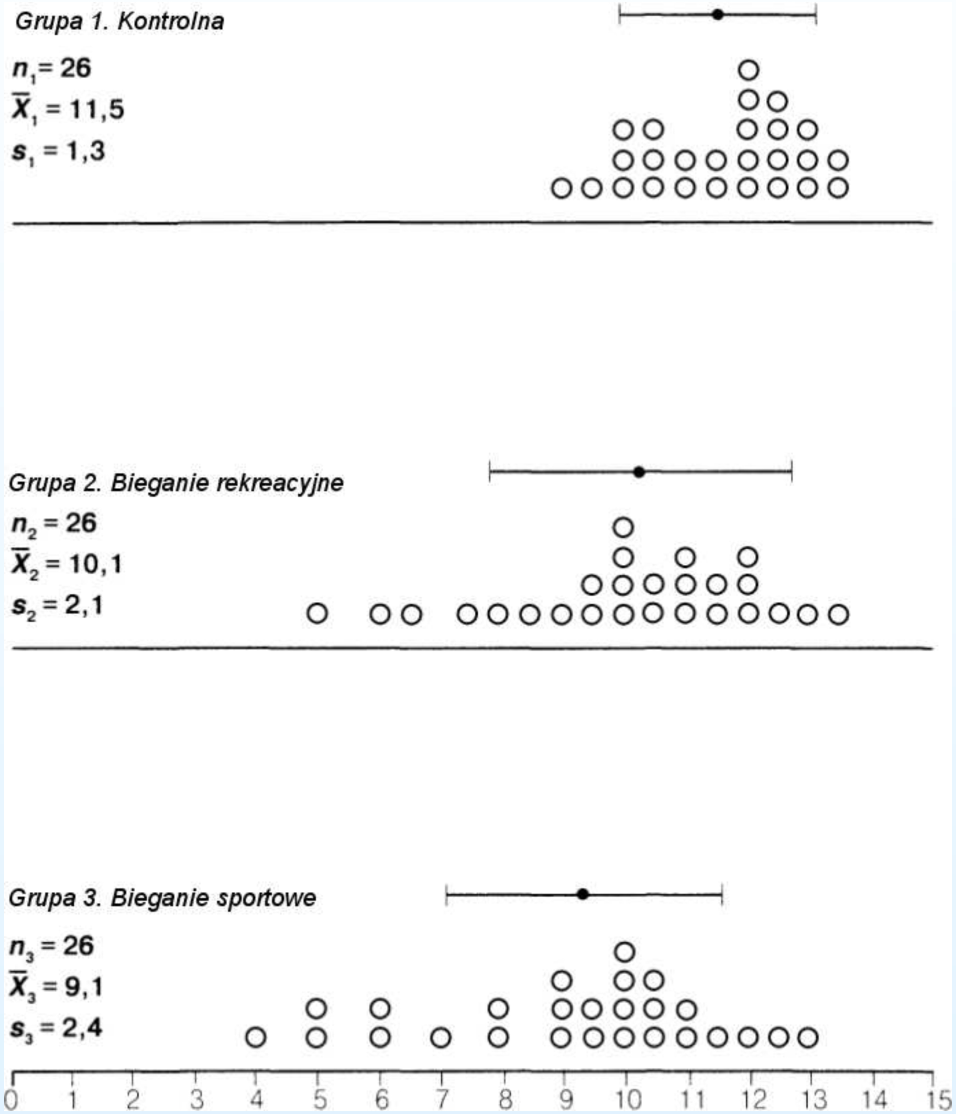
$$s^2_{\mathrm{wewn}} = \frac{1}{m}\left(s_1^2 + \cdots + s_m^2\right)$$

- $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2_{\mathrm{wewn}}}{n_1} + \frac{s^2_{\mathrm{wewn}}}{n_2}}}$,
- dla prób tego samego rozmiaru $t = \frac{\bar{X}_1 = \bar{X}_2}{\sqrt{\frac{2s^2_{\mathrm{wewn}}}{n}}}$,

Modyfikacja korekcji Bonferroniego, cd

- Ilość stopni swobody wyniesie $\nu = m(n-1)$
- Jeżeli ilość grup $m > 2$, to ilość stopni swobody będzie większa niż $2(n-1)$
- Co powoduje zmniejszenie wartości krytycznej t

Sport a zaburzenia miesiączkowania



Sport a zaburzenia miesiączkowania. Wyniki

- $s^2_{\text{wewn}} = 3{,}95$
- $m = 3, n = 26 \Rightarrow \nu = 75$
- $t_{2,1} = -2{,}54$
- $t_{3,1} = -4{,}35$
- $t_{2,3} = 1{,}81$
- poziom ufności $\alpha = 0{,}05/3 \approx 0,017$
- $t_{\text{krit}} = 2{,}45$
- wnoski:
  - bieganie powoduje zmniejszenie częstotliwości miesiączkowania
  - nie ma różnicy statystycznej między bieganiem sportowym a bieganiem rekreacyjnym

Test Newmana-Keulsa

- Bardziej dokładne oszacowanie α' niż korekta Bonferroniego, zwłaszcza dla dużej ilości porównań
  - Za pomocą analizy wariancji sprawdzamy hipotezę zerową
  - Jeżeli zostaje odrzucona, to

1. uporządkujemy grupy rosnąco względem średnich
2. dla wszystkich par obliczmay test Newmana-Keulsa

$$q = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s^2_{\mathrm{wewn}}}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

3. porównujemy wynik z wartością ktytyczną z tabeli

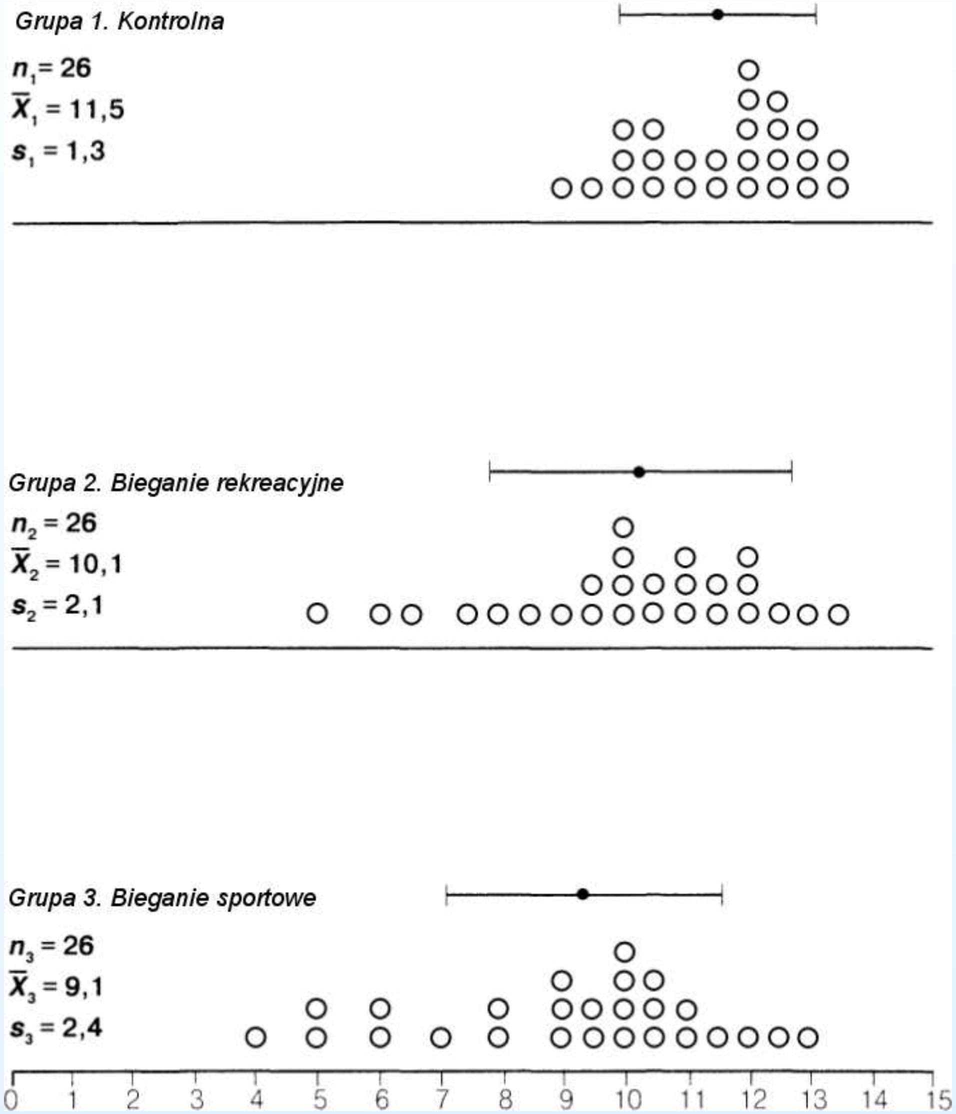
Wartośś krytyczna testu Newmana-Keulsa

- Zależy od
  - α' — prawdopodobieństwo pomylić się przynajmniej w jednym porównywaniu
  - ilość stopni swobody $\nu = N - m$, gdzie N jest sumą liczebności wszystkich grup, a m jest ilością grup
  - przedział porównywań l
    - jeżeli porównujemy grupę j z grupą i, to $l = j - i + 1$

Kolejność porównywania testu Newmana-Keulsa

- Grupy uporządkowane rosnąco: 1 — minimalna średnia, m — maksymalna
  - $m \leftrightarrow 1, m \leftrightarrow 2, \ldots, m \leftrightarrow (m-1)$
  - $(m-1) \leftrightarrow 1, (m-1) \leftrightarrow 2, \ldots, (m-1) \leftrightarrow (m-2)$
  - $\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
  - $3 \leftrightarrow 1, 3 \leftrightarrow 2$
  - $2 \leftrightarrow 1$
- grupy i i j statystycznie się nie różnią, to nie ma potrzeby porównywać wszystkich pośrednich par
  - przykładowo, jeżeli grupy 3 i 1 się nie różnią, to się nie różnią grupy 3 i 2, ani 2 i 1

Sport a zaburzenia miesiączkowania



Analiza

- Uporządkujemy rosnąco
  1. Bieganie sportowe $\bar{X}_1 = 9{,}1$
  2. Bieganie rekreacyjne $\bar{X}_2 = 10{,}1$
  3. Grupa kontrolna $\bar{X}_3 = 11{,}5$
- $s_{\text{wewn}} = 3{,}95$, $\nu = 75$, $n_1 = n_2 = n_3 = 26$
- $3 \leftrightarrow 1$
  - $q = 6{,}157$
  - $l = 3$, wartość krytyczna dla $\alpha' = 0{,}05$ wynosi $3{,}385$
- $3 \leftrightarrow 2$
  - $q = 3{,}592$
  - $l = 2$, wartość krytyczna dla $\alpha' = 0{,}05$ wynosi $2{,}822$
- $2 \leftrightarrow 1$
  - $q = 2{,}566$

Test Tukeya

- Jak test Newmana-Keulsa, tylko zamiast l zawsze się bierze m
- Czyli wartość krytyczna jest taka sama
- W poprzednim przykładzie $3{,}385$
  - $3 \leftrightarrow 1$
    - $q = 6{,}157$
  - $3 \leftrightarrow 2$
    - $q = 3{,}592$
  - $2 \leftrightarrow 1$
    - $q = 2{,}566$
- jest bardziej restrykcyjny, częsciej akceptuje hipotezę zerową (częsciej odrzuca istnienie różnicy statystycznej)

Porównywanie z grupą kontrolną

- Zagadnienie polega na porównywaniu kilka grup z kontrolną
- Poprzednie metody mogą niepotrzebnie zaostrzyć kriteria z powodu dużej ilości zbędnych porównywań
- Stosujue się inne, specjalnie opracowane metody
- Pierwszym krokiem zawsze jest analiza wariancji
  - inne metody porównywania używamy tylko jeżeli hipoteza zerowa zostaje odrzucona

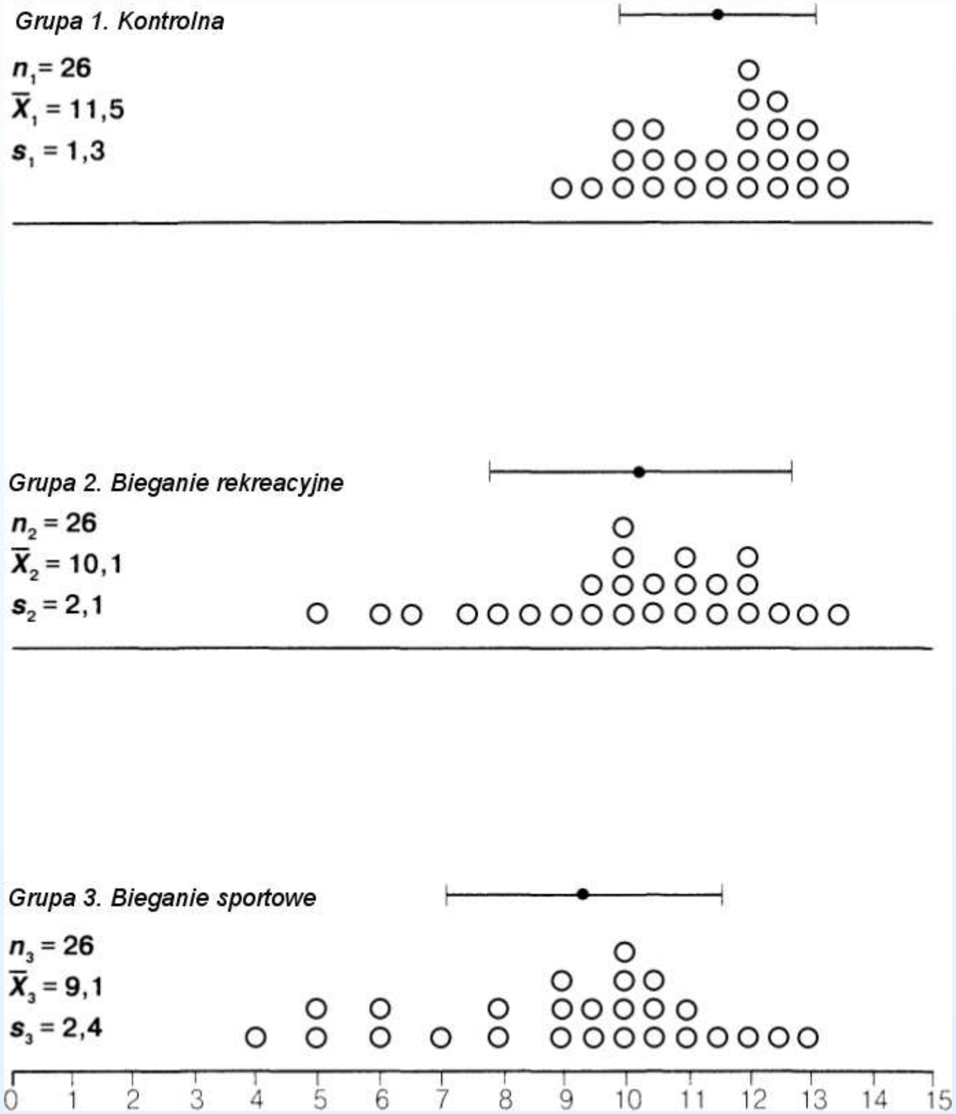
Korekta Bonferroniego

- Jak wyżej, z tym że ilość porównywań jest mniejsza
- $\alpha' = \alpha/k$
- W poprzednim przykładzie:
    - trzy grupy, jedna kontrolna $\Rightarrow$ 2 porównywania
    - $\alpha' = 0{,}05/2 = 0{,}025$ (było $0{,}017$)
    - $\nu = 75$
    - wartość krytyczna $t = 2{,}31$ (było $2{,}45$)
        - $t_{1,2} = 2{,}54$
        - $t_{1,3} = 4{,}35$
    - nic nie stwierdzamy o różnicy biegu rekreacyjnego i sportowego

Test Dunnetta

- Jest odmianą testu Newmana-Keulsa
- $q' = \dfrac{\bar{X}_{\mathrm{kontr}} - \bar{X}_A}{\sqrt{s^2_{\mathrm{wewn}}\left(\frac{1}{n_{\mathrm{kontr}}} + \frac{1}{n_A}\right)}}$
- Ilość porównań jest ilością grup (bez kontrolnej)
  - a więc mniejsza jest wartość krytyczna
- Upodrządkowanie według wartości bezwzględnej różnicy średniej od średniej grupy kontrolnej
- Porównywanie się zaczyna od najwięcej różniącej się grupy
- Jeżeli pewna grupa nie różni się statystycznie od kontrolnej, to porównywanie się kończy
- Parametr l zgadza z ilością grup m (łącznie z kontrolną)
- $\nu = N - m$ (jak w teście Newmana-Keulsa)

Sport a zaburzenia miesiączkowania



Analiza

- Bieganie sportowe:
  - $q' = 4{,}35$
  - $\nu = 75$
  - $l = 3$, wartość krytyczna dla $\alpha' = 0{,}05$ wynosi $2{,}26$
- Bieganie rekreacyjne $\bar{X}_2 = 10{,}1$
  - $q' = 2{,}54$

$P < 0{,}05$

- P jest prawdopodobieństwem omyłkowo odrzucić hipotezę zerową (znaleźć różnicę tam, gdzie jej nie ma)
- P jest prawdopodobieństwem błędu pierwszego rodzaju
- błąd drugiego rodzaju — omyłkowo przyjąć hipotezę zerową (nie zauważyć istniejącej różnicy)
- P nic nie mówi o błędzie drugiego rodzaju