Statystyka medyczna

Krzysztof S. Gołba

Klinika Elektrokardiologii i Niewydolności Serca
Oddział Elektrokardiologii
Górnośląskie Centrum Medyczne
Śląskiego Uniwersytetu Medycznego w Katowicach

Szukaj wiedzy

MEDCALC®
easy-to-use statistical software

MedCalc.org

FB MedCalc Software

UWAGA ! Informacja o zaliczeniu na ostatnim slajdzie

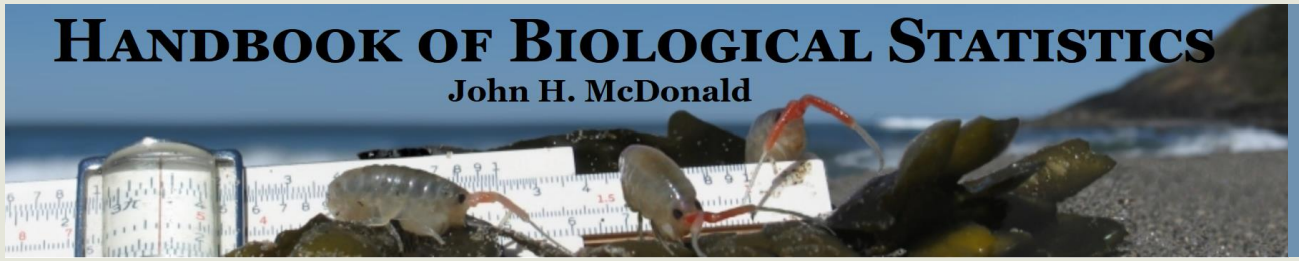
Szukaj wiedzy

Naukowiec.org

https://www.naukowiec.org/wzory/statystyka.html
https://www.naukowiec.org/tablice/statystyka.html
https://www.naukowiec.org/wiedza/statystyka.html
https://www.naukowiec.org/narzedzia/statystyka.html

StatSoft Electronic Statistics Textbook

StatSoft Podręcznik PL



Biostat Handbook

Podstawowe pojęcia

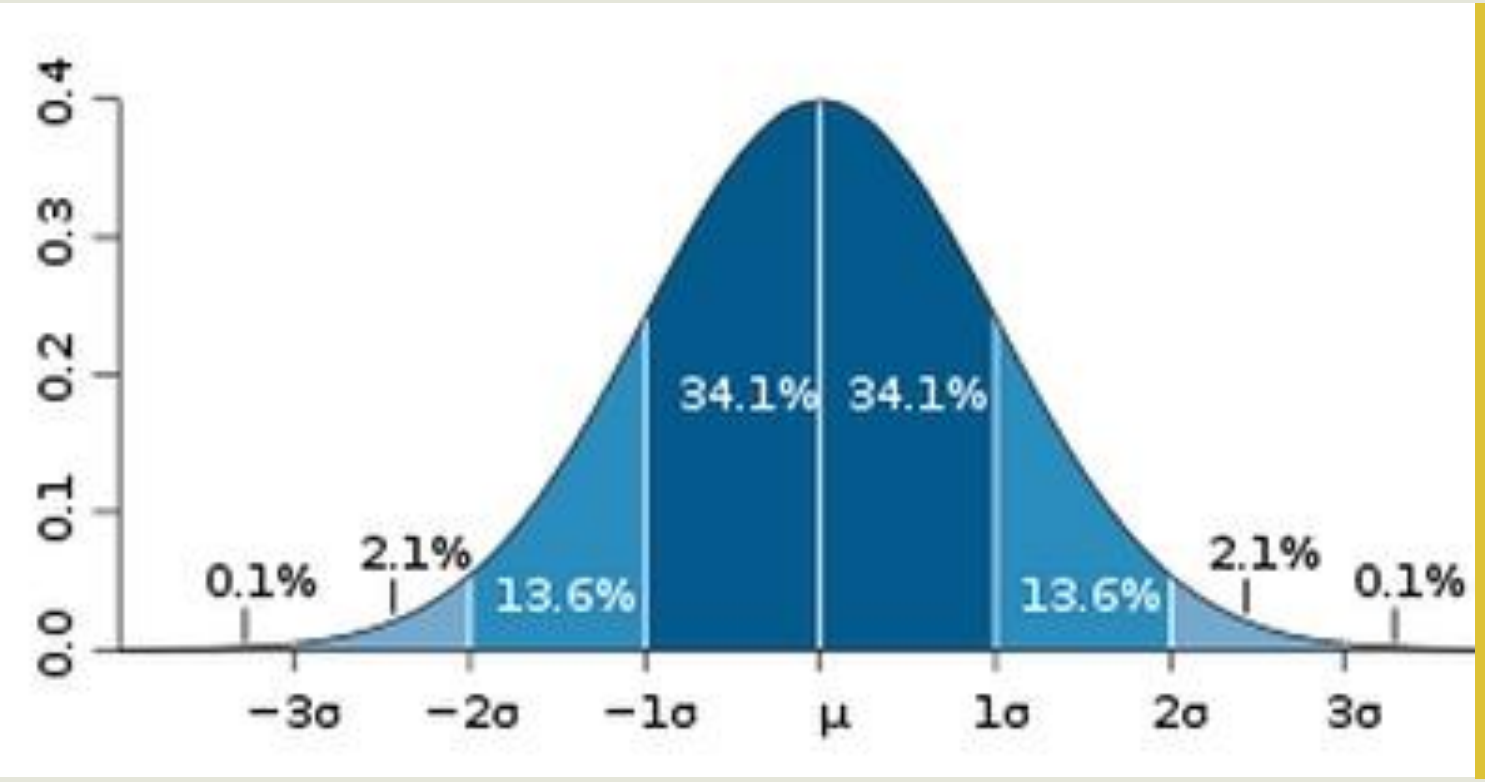
Zmienne są to wielkości, które rejestrujemy, mierzymy, kontrolujemy lub ustalamy w trakcie prowadzenia badań.

Istotność statystyczna różnicy zmiennych. Porównując ze sobą grupy w aspekcie określonej zmiennej zakładamy, że istnieje różnica pomiędzy grupami w zakresie tej zmiennej. Musimy zatem zastosować właściwy test statystyczny. Zanim go przeprowadzimy należy przyjąć określony poziom istotności różnic, który pozwoli uznać że wyniki są istotnie statystycznie różne. Gdy przyjmiemy poziom istotności (prawdopodobieństwo, ze różnica jest przypadkowa) $p < 0,05$ to zaobserwowane różnica, dla których p jest mniejsze niż 0,05 możemy nazwać różnicą istotną statystycznie.

Czytaj dalej Podstawowe pojecia

Czytaj dalej Wyniki istotne staystycznie

Rozkład normalny - Rozkład Gaussa

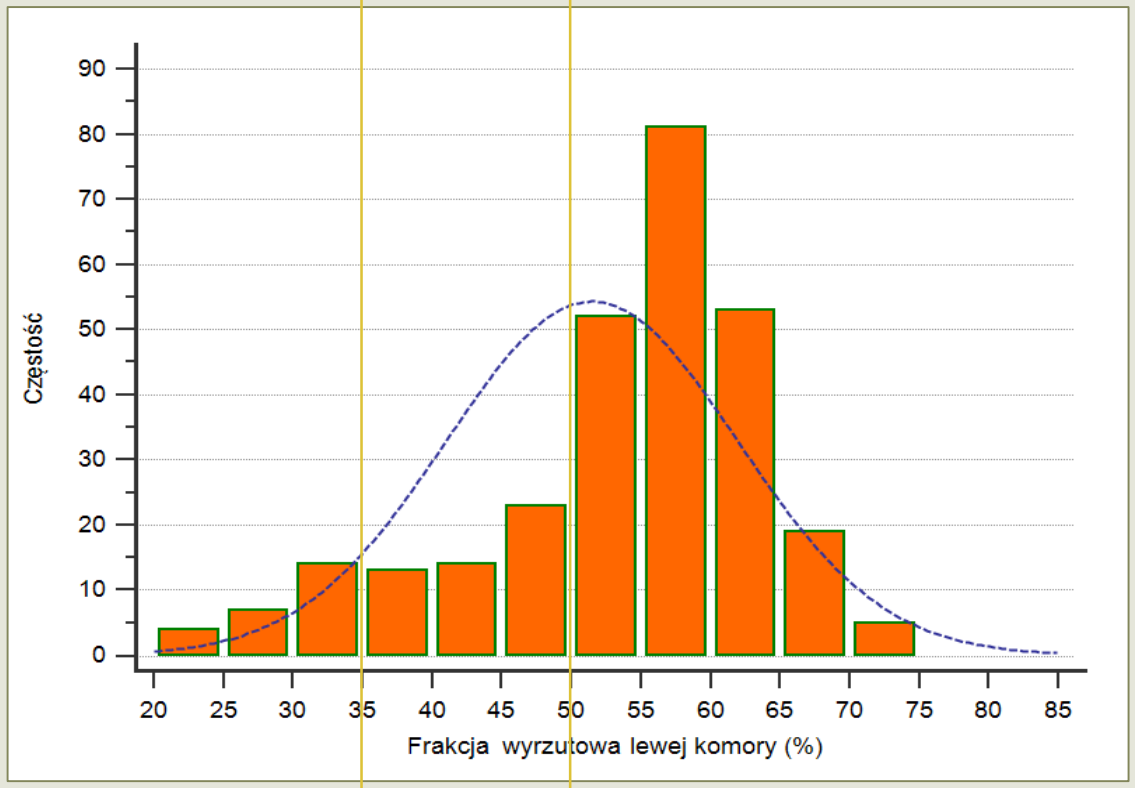


Rozkład normalny, inaczej zwany rozkładem Gaussa, krzywą Gaussa jest najważniejszym rozkładem teoretycznym prawdopodobieństwa w statystyce. Rozkład normalny jest też najbardziej intuicyjnym rozkładem statystycznym. Opisuje on sytuacje w świecie, gdzie większość przypadków jest bliska średniemu wynikowi, a im dany wynik bardziej odchyla się od średniej tym jest mniej reprezentowany.

- Najwięcej jest przypadków blisko przeciętnej.
- Im dalej oddalamy się od średniego wyniku, tym przypadków jest mniej.

Czytaj dalej

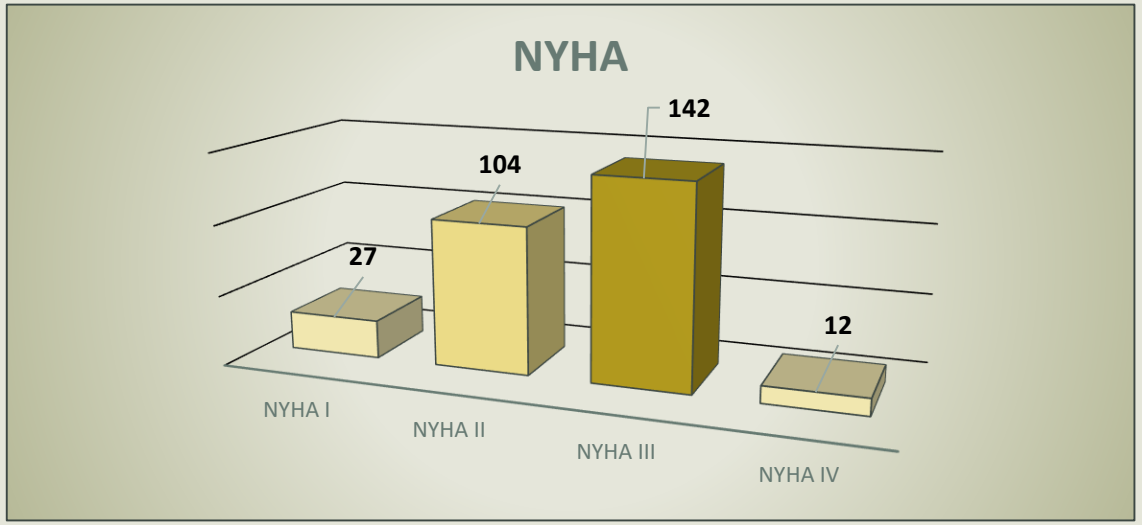
Rozkłady danych – real life



Celem pracy była ocena czynników predykcyjnych konieczności zastosowania stałej stymulacji serca (PPI) u pacjentów poddanych przezcewnikowej implantacji zastawki aortalnej (TAVI) we wczesnym okresie pooperacyjnym do 30 doby włącznie po zabiegu.

Rozkład frakcji wyrzutowej lewej komory (LVEF) z medianą 55 (54-55) %

Rozkłady danych – real life



Celem pracy była ocena czynników predykcyjnych konieczności zastosowania stałej stymulacji serca (PPI) u pacjentów poddanych przezcewnikowej implantacji zastawki aortalnej (TAVI) we wczesnym okresie pooperacyjnym do 30 doby włącznie po zabiegu.

Rozkład pacjentów z określoną klasą czynnościowa wg klasyfikacji NYHA (New York Heart Association).

Testy parametryczne vs. nieparametryczne

Testy parametryczne
trzeba spełnienia więcej założeń
Ale
testy mają większą moc
(trudniej odrzucić H0)
dokładniejszy pomiar
łatwiej interpretować wyniki

Testy nieparametryczne
można spełnienić mniej założeń
Ale
testy mają mniejszą moc
(łatwiej odrzucić H0)
mniej dokładny pomiar
trudniej interpretować wyniki

Podstawowe warunki do zastosowania testu parametrycznego

rozkład normalny badanych zmiennych	zmienne muszą być mierzone na skali ilościowej

Testy parametryczne
testy t-Studenta
analiza wariancji
korelacja r-Pearsona
analiza regresji

Testy nieparametryczne
test U Manna-Whitneya
test Kruskala-Wallisa
test niezależności chi^2 (Pearsona)
korelacja tau-b Kendalla

Miary położenia w statystyce

Miary położenia

Miary położenia wskazują wokół jakich wartości skupia się rozkład analizowanych zmiennych. Dzieli się je na **klasyczne** i **pozycyjne**. Miary klasyczne to średnie: arytmetyczna, harmoniczna i geometryczna. Do miar pozycyjnych należy dominanta (modalna, wartość najczęstsza) oraz kwantyle. Wśród kwantyli najczęściej stosowane są: kwartyle (dzielące zbiorowość na cztery części pod względem liczebności), kwintyle (dzielące zbiorowość na pięć części), decyle (dzielące zbiorowość na dziesięć części) oraz percentyle (dzielące zbiorowość na sto części).

Obydwie grupy średnich (klasyczne i pozycyjne) nie wykluczają się, ale nawzajem uzupełniają. Każda z nich opisuje bowiem poziom wartości cechy z innego punktu widzenia.

Literatura: Miary położenia / eks.stat.gov.pl / GUS

Średnia arytmetyczna

Jest ilorazem sumy wartości zmiennej i liczebności badanej zbiorowości. Z szeregów wyliczających średnią arytmetyczną obliczamy następująco:

$$\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{N} = \frac{\sum_{j=1}^{N} x_j}{N},$$

gdzie:

$\bar{x}$ – symbol średniej arytmetycznej,

x_j – warianty cechy mierzalnej,

N – liczebność badanej zbiorowości.

Mediana (Kwantyle)

Kwantyle są to takie wartości cechy, które dzielą badaną zbiorowość na określone, części pod względem liczebności. Spośród kwantyli najczęściej używanymi miarami są **kwartyle**, wśród których wyróżniamy kwartyl pierwszy (dolny), kwartyl drugi (mediana) oraz kwartyl trzeci (górny). Każdy z kwartyli dzieli zbiorowość uporządkowaną niemalejąco pod względem liczebności na dwie części, przy czym:

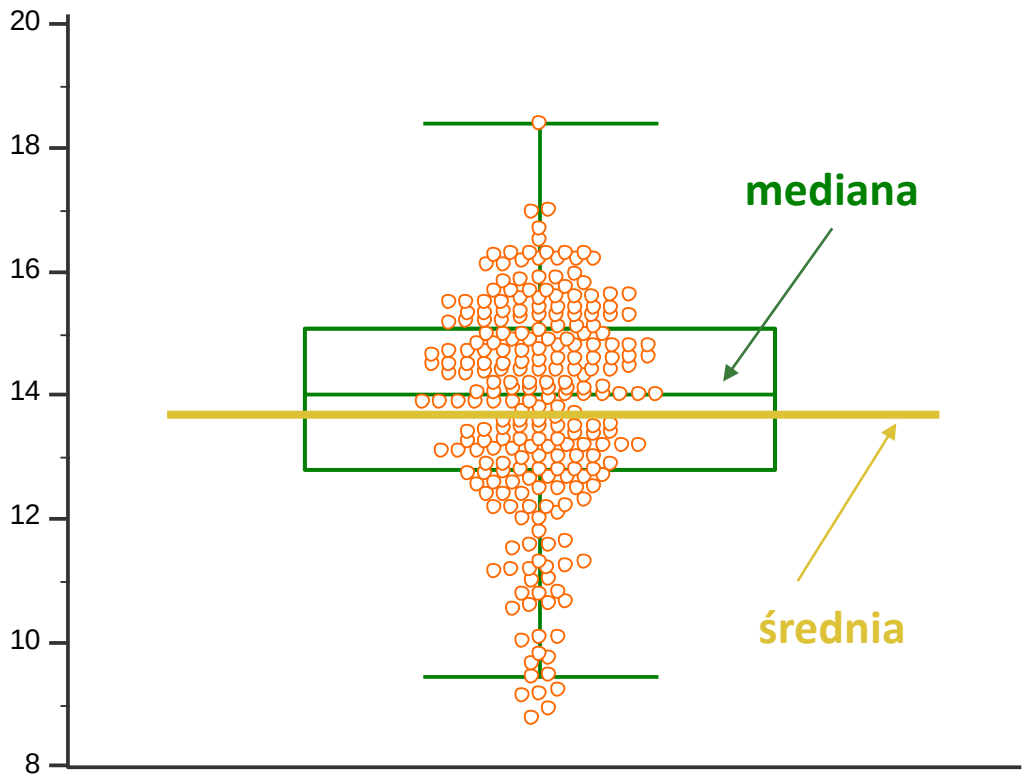
- **kwartyl pierwszy** dzieli zbiorowość na dwie części w ten sposób, że 25% jednostek zbiorowości ma wartości zmiennej mniejsze lub równe kwartylowi pierwszemu, a 75% - równe lub większe od tego kwartyla,
- **mediana** dzieli zbiorowość na dwie części w ten sposób, że 50% jednostek ma wartości mniejsze lub równe medianie oraz 50% - równe lub większe od mediany,
- **kwartyl trzeci** dzieli zbiorowość na dwie części w ten sposób, że 75% jednostek ma wartości zmiennej mniejsze lub równe kwartylowi trzeciemu, a 25% - równe lub większe od kwartyla trzeciego.

Miary położenia w statystyce

Średnia vs. mediana

Zmienna	HEMOGLOBINA	(stężenie hemoglobiny g/dl)

Liczność próby	252
Najniższa wartość	8,8000
Najwyższa wartość	18,4000
Średnia arytmetyczna	13,7885
95% CI dla średniej arytmetycznej	13,5718 do 14,0052
Mediana	14,0300
95% CI dla mediany	13,7971 do 14,3508
Wariancja	3,0518
Standardowe odchylenie	1,7469
Względne standardowe odchylenie	0,1267 (12,67%)
Standardowy błąd średniej	0,1100
Współczynnik skośności	-0,6588 (P=0,0001)
Współczynnik kurtozy	0,2652 (P=0,3501)



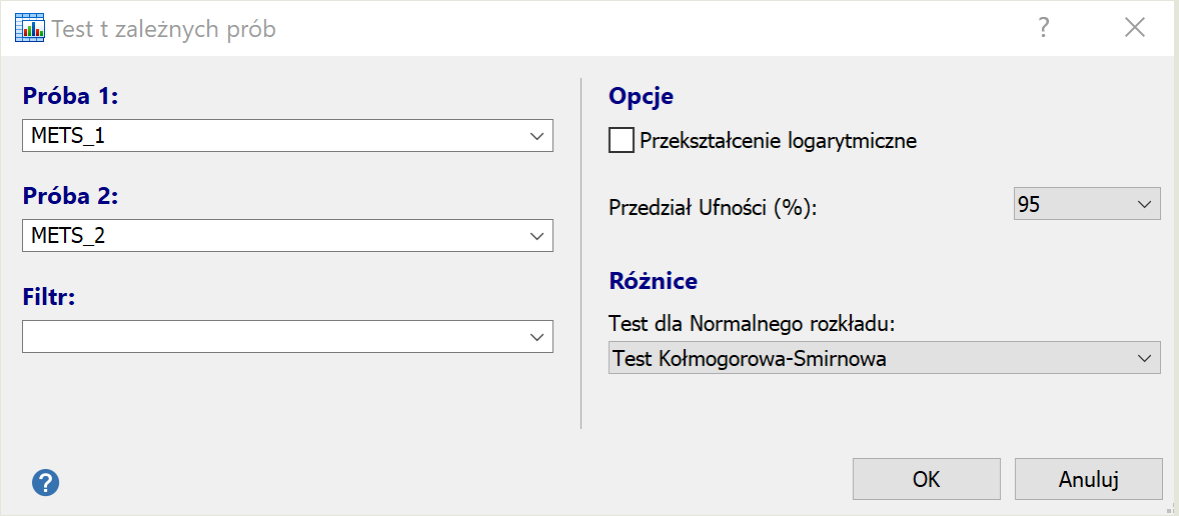
Statystyki opisowe - czytaj dalej

Test t Studenta dla prób zależnych

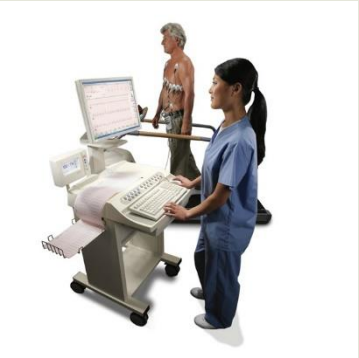
Najczęściej stosowana metoda statystyczna do porównania średnich z dwóch zależnych od siebie grup np. ci sami pacjenci przed i po leczeniu

283 pacjentów po angioplastyce wieńcowej rehabilitowano w warunkach oddziału dziennego rehabilitacji w ramach II etapu wczesnej rehabilitacji. Przed i po cyklu rehabilitacji wykonywano u każdego pacjenta próbę wysiłkową na bieżni ruchomej. Wynik próby określał submaksymalną wydolność krążeniowo-oddechową (wydolność fizyczną) jednego konkretnego pacjenta w MET: przed METS_1 vs. METS_2.

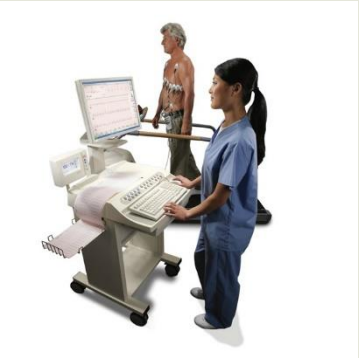
Test t Studenta dla prób zależnych porównuje dwie średnie parametrów pochodzących od tych samych osoby przed i po np. leczeniu. Hipoteza zerowa H0: rehabilitacja nie zmieniła wydolności fizycznej pacjentów, Hipoteza alternatywna H1 rehabilitacja kardiologiczna poprawia wydolność sercowo naczyniową



METS_1



METS_2



Test t Studenta dla prób zależnych

Najczęściej stosowana metoda statystyczna do porównania średnich z dwóch zależnych od siebie grup np. ci sami pacjenci przed i po leczeniu

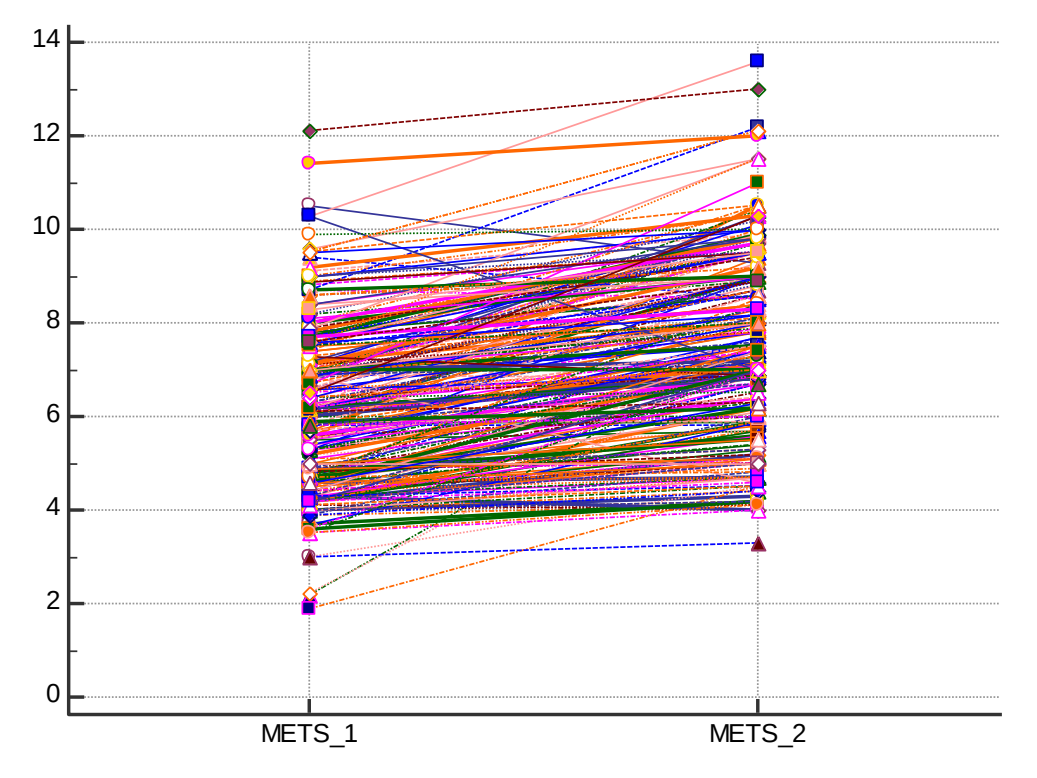
E	F
METS_1	METS_2
6	7
6	9
6	11
5	7
6	6
6	8
5	8
9	10
6	8
8	10
5	5
7	8
5	6
5	5
6	6
6	9
7	10
6	7
7	8
6	6
6	7
7	8
6	7
7	9
7	8
5	8
6	8
5	5
6	8
6	7
9	9
5	6
6	7
5	5
4	6
4	6
5	7
7	7
5	6
7	8
7	7
8	9
5	5
6	7
2	8
5	7
7	8

Próba 1	METS_1
Próba 2	METS_2

	Próba 1	Próba 2
Liczność próby	283	283
Średnia arytmetyczna	6,1199	7,4788
95% CI dla średniej	5,9211 do 6,3188	7,2485 do 7,7091
Wariancja	2,8887	3,8731
Standardowe odchylenie	1,6996	1,9680
Standardowy błąd średniej	0,1010	0,1170

Test t zależnych prób

Średnia różnica	1,3589
Standardowe odchylenie średniej różnicy	1,1533
Standardowy błąd średniej różnicy	0,06856
95% CI różnicy	1,2239 do 1,4938
Test statystyki t	19,820
Stopnie swobody (DF)	282
Dwustronne prawdopodobieństwo	P < 0,0001



Wydolność fizyczna pacjentów w wyniku rehabilitacji poprawiła się istotnie, średnio o 1,3±1,15 MET

Paired samples t-test

Description

The paired t-test is used to test the null hypothesis that the average of the differences between a series of paired observations is zero. Observations are paired when, for example, they are performed on the same samples or subjects.

Required input

Identify the variables for sample 1 and sample 2. You can use the button to select variables and filters.

Results

The results windows for the paired samples t-test displays the summary statistics of the two samples. Note that the sample size will always be equal because only cases are included with data available for the two variables.

Next, the arithmetic mean of the differences (*mean difference*) between the paired observations is given, the standard deviation of these differences and the standard error of the mean difference followed by the confidence interval for the mean difference.

Differences are calculated as *sample 2 – sample 1*.

In the paired samples t-test the null hypothesis is that the average of the differences between the paired observations in the two samples is zero.

If the calculated P-value is less than 0.05, the conclusion is that, statistically, the mean difference between the paired observations is significantly different from 0.

Note that in MedCalc P-values are always two-sided.

Normal distribution of differences

For the paired samples t-test, it is assumed that the differences of the data pairs follow a Normal distribution (you do not need to check for normality in the two separate samples). This assumption can be evaluated with a formal test, or by means of graphical methods.

The different formal Tests for Normal distribution may not have enough power to detect deviation from the Normal distribution when sample size is small. On the other hand, when sample size is large, the requirement of a Normal distribution is less stringent because of the central limit theorem.

Therefore, it is often preferred to visually evaluate the symmetry and peakedness of the distribution of the differences using the Histogram, Box-and-whisker plot, or Normal plot.

To do so, you click the hyperlink "Save differences" in the results window. This will save the differences between the paired observations as a new variable in the spreadsheet. You can then use this new variable in the different distribution plots.

Literature

- Altman DG (1991) Practical statistics for medical research. London: Chapman and Hall.

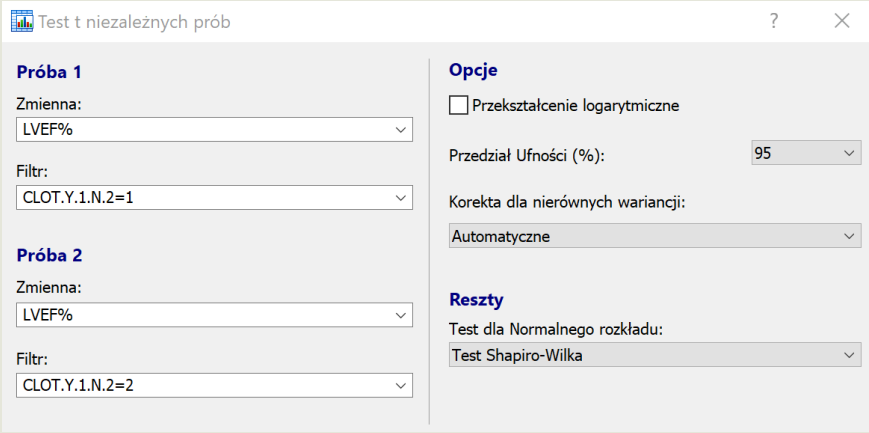
Testy t Studenta

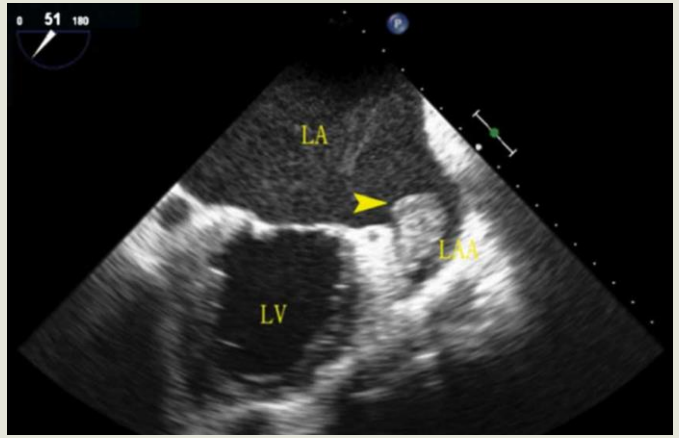
Testy t Studenta

Test t Studenta dla prób niezależnych

Najczęściej stosowana metoda statystyczna do porównania średnich z dwóch niezależnych od siebie grup

350 pacjentów przygotowywano do kardiowersji elektrycznej. Pacjenci byli wcześniej poddani rutynowej terapii przeciwzakrzepowej. W dniu poprzedzającym kardiowersję wykonano u każdego z pacjentów badanie echokardiograficzne przezprzełykowe w celu wykluczenia obecności skrzepliny w świetle lewego przedsionka, a szczególnie wuszku lewego przedsionka. U 47 z nich stwierdzono skrzeplinę (CLOT.Y.1.N.2=1) i zrezygnowano z wykonania kardiowersji. Z kolei u 303 przedsionek i jego uszko były wolne od skrzepliny (CLOT.Y.1.N.2=2). Zadano pytanie czy jedna z przyczyn obecności skrzepliny mogła być niższa frakcja wyrzutowa lewej komory serca (LVEF%) u 47 pacjentów w porównaniu z 303 pacjentami bez skrzepliny. Test t Studenta dla prób niezależnych porównuje dwie średnie parametrów pochodzących z dwóch różnych grup jak powyższym przykładzie. Hipoteza zerowa H0: frakcja wyrzutowa lewej komory nie była różna u pacjentów z i bez skrzepliny. Hipoteza alternatywna H1: frakcja wyrzutowa lewej komory była istotnie różna u pacjentów z i bez skrzepliny.





Test t Studenta dla prób niezależnych

Najczęściej stosowana metoda statystyczna do porównania średnich z dwóch niezależnych od siebie grup

Test t niezależnych prób

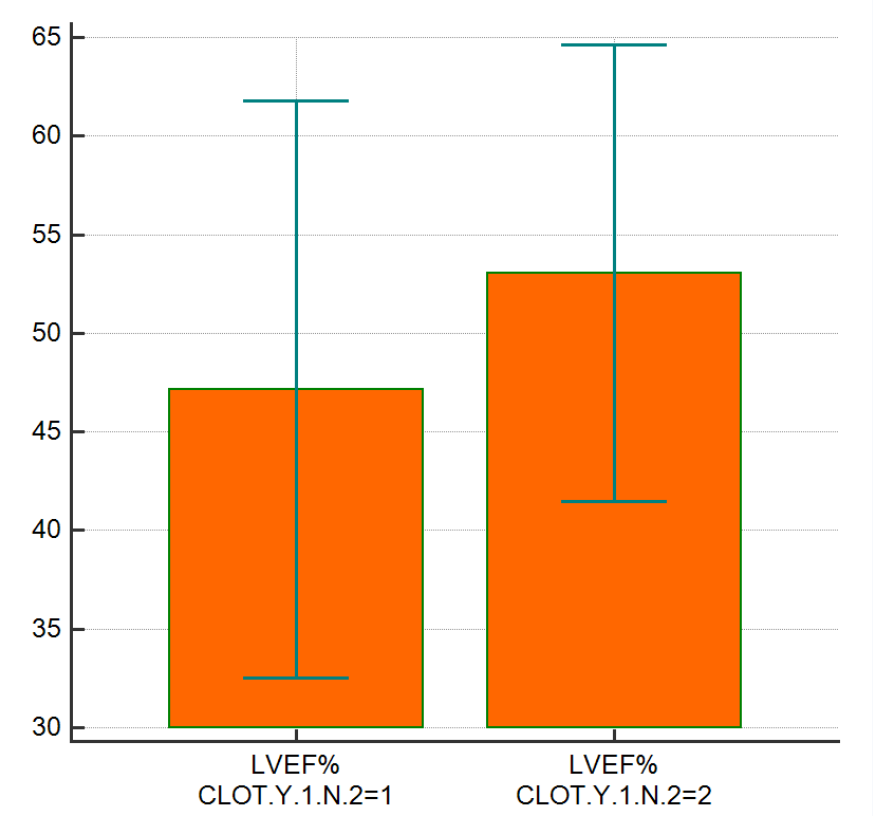
Próba 1	
Zmienna	LVEF%
Filtr	CLOT.Y.1.N.2=1
Próba 2	
Zmienna	LVEF%
Filtr	CLOT.Y.1.N.2=2

	Próba 1	Próba 2
Liczność próby	47	303
Średnia arytmetyczna	47,1702	53,0561
95% CI dla średniej	42,8745 do 51,4659	51,7473 do 54,3649
Wariancja	214,0574	134,0399
Standardowe odchylenie	14,6307	11,5776
Standardowy błąd średniej	2,1341	0,6651

Test F dla równych wariancji	P = 0,024

Test Welcha (zakładane nierówne wariancje)

Różnica	5,8859
Standardowy Błąd	2,2353
95% CI różnicy	1,4062 do 10,3656
Statystyki testu t(d)	2,633
Stopnie swobody (DF)	55,3
Dwustronne prawdopodobieństwo	P = 0,0110



Frakcja wyrzutowa lewej komory (LVEF%) była istotnie różna u pacjentów z (CLOT.Y.1.N.2=1) i bez skrzepliny (CLOT.Y.1.N.2=2), p=0,0110.

Independent samples t-test

Description

Use the Independent samples t-test to compare the mean of two independent samples. It tests the null hypothesis that the difference between the means of two samples is equal to 0.

Required input

Identify the variables for sample 1 and sample 2. Differences will be calculated as Sample2−Sample1.

Caveat: the two filters must define distinct groups so that the same case is not included in the two samples.

Results

The results windows for the Independent samples t-test displays the summary statistics of the two samples, followed by the statistical tests.

First an F-test is performed. If the P-value is low ($P<0.05$) the variances of the two samples cannot be assumed to be equal and it should be considered to use the t-test with a correction for unequal variances (Welch test, Armitage et al., 2002).

The independent samples t-test is used to test the hypothesis that the difference between the means of two samples is equal to 0 (this hypothesis is therefore called the null hypothesis). The program displays the difference between the two means, the pooled standard deviation, standard error, and the confidence interval (CI) of this difference. Next follow the test statistic t, the Degrees of Freedom (DF) and the two-tailed probability P. When the P-value is less than the conventional 0.05, the null hypothesis is rejected and the conclusion is that the two means do indeed differ significantly.

Note that in MedCalc P-values are always two-sided.

Normal distribution of residuals

For the independent samples t-test, it is assumed that the residuals (the differences between the observations and their group or sample mean) follow a Normal distribution. This assumption can be evaluated with a formal test, or by means of graphical methods.

The different formal Tests for Normal distribution may not have enough power to detect deviation from the Normal distribution when sample size is small. On the other hand, when sample size is large, the requirement of a Normal distribution is less stringent because of the central limit theorem.

Literature

- Armitage P, Berry G, Matthews JNS (2002) Statistical methods in medical research. 4th ed. Blackwell Science.
- Bland M (2000) An introduction to medical statistics, 3rd ed. Oxford: Oxford University Press.

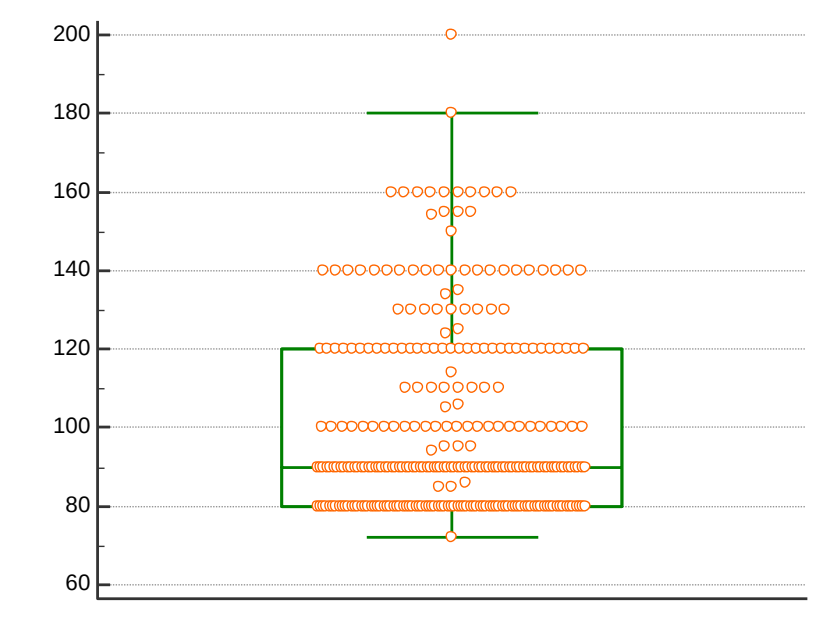
Testy t Studenta

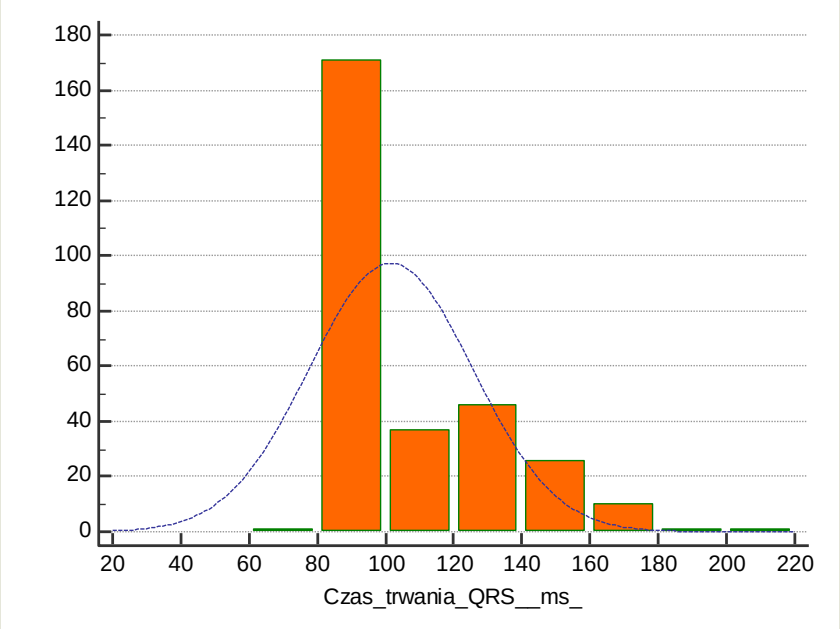
Testy t Studenta

Test U Mann-Whitney'a

Zastawkę aortalną metoda przezskórną (TAVI) wszczepiono 293 pacjentom. U 43 jako powikłanie wystąpił blok przedsionkowo-komorowy III stopnia. Konieczne było wszczepienie u tych pacjentów rozrusznika. Postanowiono sprawdzić czy czas trwania zespołu QRS w elektrokardiogramie różni się w grupach pacjentów, u których wystąpiła lub nie wystąpiła konieczność wszczepienia stymulatora.

Chciano zastosować test t Studenta dla danych niepowiązanych. Jednak okazałao się, że rozkład warosci czasu trwania QRS w całej badanej grupie nie jest normalny.





Statystyki sumaryczne

Zmienna	Czas_trwania_QRS__ms_
Liczność próby	293
Najniższa wartość	72,0000
Najwyższa wartość	200,0000
Średnia arytmetyczna	101,5666
95% CI dla średniej arytmetycznej	98,7992 do 104,3339
Mediana	90,0000
95% CI dla mediany	90,0000 do 90,0000
Wariancja	579,2875
Standardowe odchylenie	24,0684
Względne standardowe odchylenie	0,2370 (23,70%)
Standardowy błąd średniej	1,4061
Współczynnik skośności	1,1970 (P<0,0001)
Współczynnik kurtozy	0,7723 (P=0,0270)
Test Shapiro-Wilka dla Normalnego rozkładu	W=0,8287 odrzuć Normalność (P<0,0001)

Test U Mann-Whitney'a

Konieczne było zastosowanie test nieparametrycznego - odpowiednika testu t-Studenta dla prób niepowiązanych. Miarą tendencji centralnej dla tego testu jest nie średnia jak w przypadków testów t, a mediana.

Test Manna-Whitneya (próby niezależne)

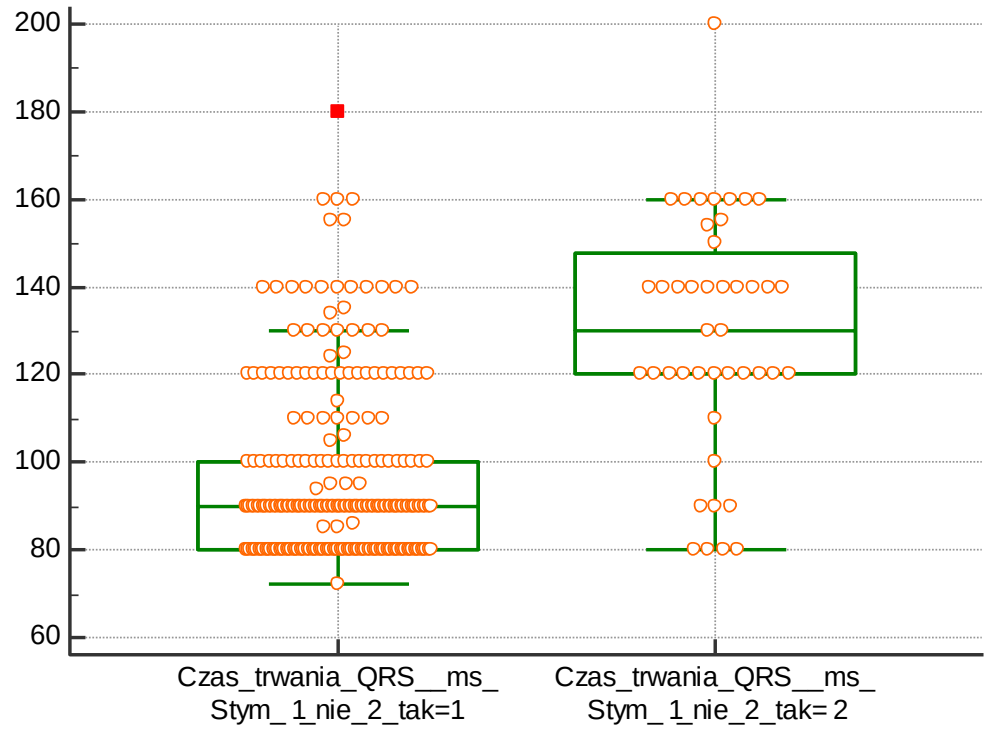
Próba 1	
Zmienna	Czas_trwania_QRS__ms_
Filtr	Wszczepienie_stymulatora__1_nie__2_tak_=1
Próba 2	
Zmienna	Czas_trwania_QRS__ms_
Filtr	Wszczepienie_stymulatora__1_nie__2_tak_=2

	Próba 1	Próba 2
Liczność próby	250	43
Najniższa wartość	72,0000	80,0000
Najwyższa wartość	180,0000	200,0000
Mediana	90,0000	130,0000
95% CI dla mediany	90,0000 do 90,0000	120,0000 do 140,0000
Rozstęp ćwiartkowy	80,0000 do 100,0000	120,0000 do 147,5000

Różnica median Hodgesa-Lehmanna	40,0000
95% Przedział Ufności	30,0000 do 40,0000

Test Manna-Whitneya (próby niezależne)

Średnia ranga pierwszej grupy	133,3620
Średnia ranga drugiej grupy	226,2907
U Manna-Whitneya	1965,50
Test statystyki Z (skorygowany dla równych wartości)	-6,804
Dwustronne prawdopodobieństwo	P < 0,0001



Grupa pacjentów wymagających wszczepienia stymulatora miała przed zabiegiem TAVI istotnie dłuższy czas trwania zespołu QRS, odpowiednio mediany (95% CI): 130 (120-140) ms v. 90 (90-90) ms, p<0,001.

Mann-Whitney test (independent samples)

Description

The Mann-Whitney test is used to test the significance of the difference between two independent samples. It is the alternative for the Independent samples t-test, when the distribution of the samples is not Normal. The Mann-Whitney test combines and ranks the data from sample 1 and sample 2 and calculates a statistic on the difference between the sum of the ranks of sample 1 and sample 2. If the resulting P-value is small (P<0.05) then a statistically significant difference between the two samples can be accepted.

Required input

Identify the variables for sample 1 and sample 2. You can use the button to select variables and filters. Caveat: the two filters must define distinct groups so that the same case is not included in the two samples.

Results

Summary statistics

The results windows for the Mann-Whitney test (independent samples) displays summary statistics of the two samples.

The statistics include the Hodges-Lehmann median difference (the Hodges-Lehmann estimate of location shift) and its 95% confidence interval (Conover, 1999). For two independent samples with sample size *m* and *n*, the Hodges-Lehmann median difference is the median of all $m \times n$ paired differences between the observations in the two samples. Differences are calculated as sample 2 – sample 1. The confidence interval is derived according to Conover (1999, p. 281).

Note that the Hodges-Lehmann median difference is not necessarily the same as the difference between the medians.

Mann-Whitney test results

The Mann-Whitney test combines and ranks the data from sample 1 and sample 2 and calculates a statistic on the difference between the sum of the ranks of sample 1 and sample 2.

In the presence of ties, MedCalc uses the method by Conover, 1999.

When there are not ties, then when either or both sample sizes are large (>20) then MedCalc uses the Normal approximation (Lentner, 1982) to calculate the P-value, and for small sample sizes MedCalc calculates the exact probability (Conover, 1999).

If the resulting P-value is small (P<0.05) then a statistically significant difference between the two samples can be accepted.

Note that in MedCalc P-values are always two-sided.

Literature

- Conover WJ (1999) Practical nonparametric statistics, 3rd edition. New York: John Wiley & Sons.
- Lentner C (ed) (1982) Geigy Scientific Tables, 8th edition, Volume 2. Basle: Ciba-Geigy Limited.

Test U Manna-Whitneya

Analiza regresji liniowej (pojedyncza)

Analiza regresji liniowej (pojedyncza, jednoczynnikowa) sprawdza czy istnieje związek pomiędzy zmienną (czynnikiem) X a zmienną zależną od X zmienną (efektem) Y.

X może zatem być nazwana zmienna wyjaśniająca lub predykcyjną –pozwalającą przewidzieć.
Analiza regresji może pozwolić sprawdzić czy efekt rehabilitacji (Y) zależy od wieku pacjent (X) lub czasu oczekiwania na rehabilitację (X).

W przypadku pojedynczej zmiennej X badamy pary XY i możemy uznać, że opisuje związek między parą zmiennych X i Y jak w analizie korelacji. Jednak korelacja zasadniczo nie musi zakładać związku przyczynowo - skutkowego. A myśląc o w analizie regresji zakładamy, że jest to związek przyczynowy.

Sprawdzimy zależność między wiekiem pacjenta **X**, a różnicą pomiędzy submaksymalna wydolność krążeniowo-oddechową (wydolność fizyczną) w drugiej w porównaniu z pierwszą próbą wysiłkową **Y** .

Szczegóły badania opisano w poprzednim przezroczu. Oznaczenia jak na wcześniejszych przeźroczach - *METS_1-METS_2 = Zmiana_METs / ΔMET.*

Analiza regresji liniowej (pojedyncza)

B	C	G
LP1	Wiek_w_latach	Zmiana_METs
1	60	0,6
2	56	3,0
3	46	5,0
4	61	2,0
5	71	0,4
6	50	2,1
7	60	2,3
8	46	0,8
9	49	1,7
10	55	1,7
11	70	-0,1
12	61	1,1
13	70	1,1
14	65	0,7
15	65	0,0
16	53	2,4
17	43	2,9
18	57	1,2
19	61	1,5
20	58	0,2
21	59	1,4
22	52	0,6
23	34	1,2
24	59	2,0
25	64	0,8
26	72	2,8
27	46	1,6
28	61	0,8
29	49	1,7
30	59	1,0
31	38	-0,9
32	70	0,5
33	72	0,9
34	76	0,1
35	61	1,7
36	57	2,0

Dependent Y	Zmiana_METs
Independent X	Wiek_w_latach
Weights	*** AutoWeight 1/SD^2 ***

Weighted least squares regression

Sample size	283
Coefficient of determination R^2	0,08642
Residual standard deviation	1,2938

Regression Equation

y = 3,2231 + -0,03179 x

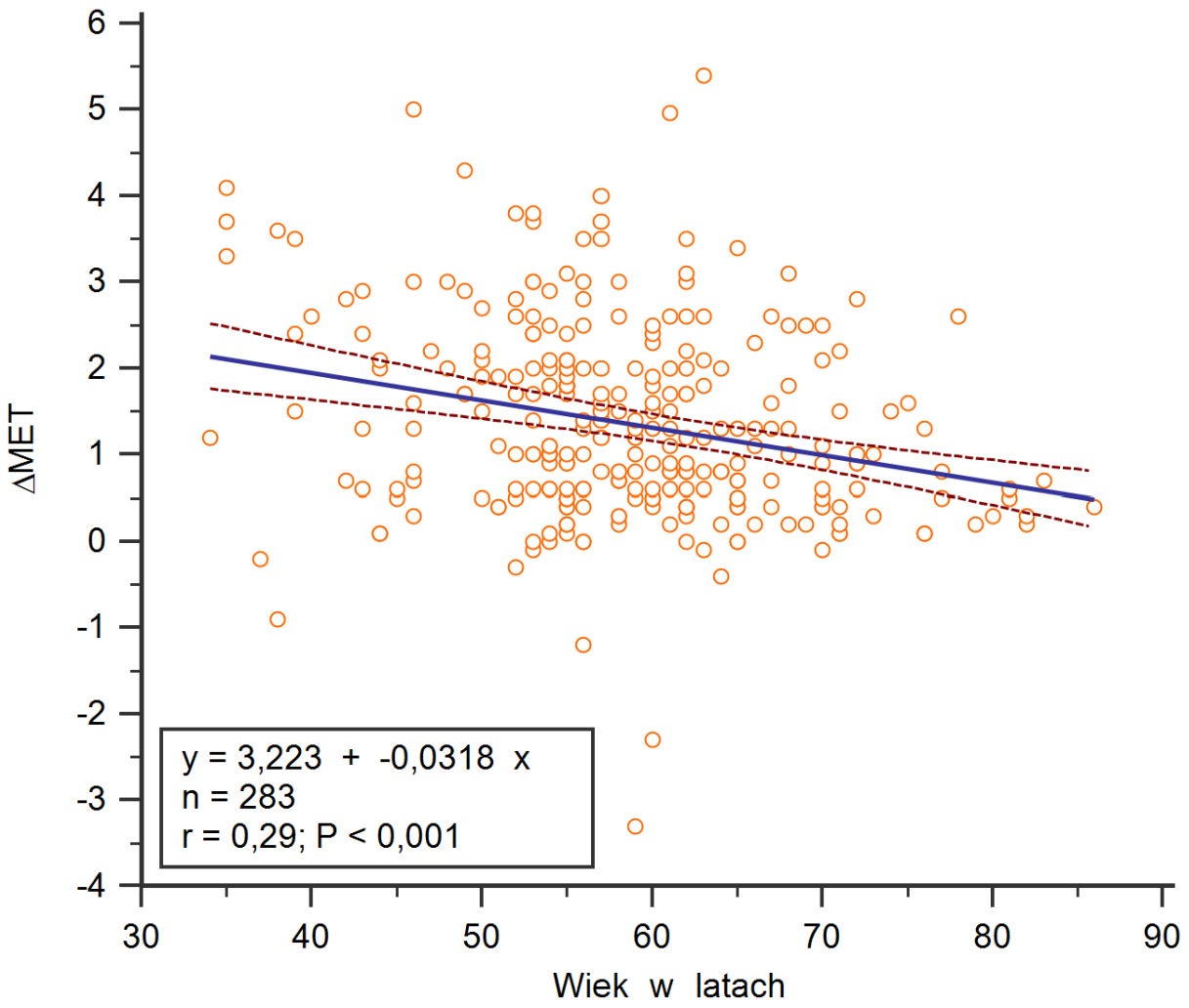
Parameter	Coefficient	Std. Error	95% CI	t	P
Intercept	3,2231	0,3918	2,4468 to 3,9994	8,2270	<0,0001
Slope	-0,03179	0,006167	-0,04393 to -0,01966	-5,1556	<0,0001

Analysis of Variance

Source	DF	Sum of Squares	Mean Square
Regression	1	44,4925	44,4925
Residual	281	470,3578	1,6739

F-ratio	26,5806
Significance level	P<0,0001

Faktycznie, ewentualna poprawa w wydolności fizycznej w wyniku rehabilitacji zależy od wieku pacjenta.
Starsi pacjenci mają tym samym gorszy efekt rehabilitacji – nie zawsze, ale jak widać zazwyczaj. Zależność jest odwrotnie proporcjonalna.



Regresja logistyczna wieloczynnikowa

	Analiza jednoczynnikowa			Analiza Wieloczynnikowa		
Parametr	**Iloraz szans**	**95%CI**	**P**	**Iloraz szans**	**95%CI**	**P**
Wiek	1,05	1-1,11	**0,05**			
Nitraty	4,74	1,56-14,49	**0,006**			
Amiodaron	3,85	1,19-12,39	**0,02**	17,71	2,48-126,67	**0,004**
LBBB	5,29	2,58-10,06	**<0,0001**			
RBBB	8,5	3,64-19,86	**<0,0001**	4,99	1,14-21,84	**0,03**
QRS (ms)	1,05	1,04-1,07	**<0,0001**	1,06	1,03-1,07	**<0,0001**
PQ (ms)	1,02	1,01-1,03	**0,004**			
Głębokość protezy w LVOT (mm)	3,3	2,16-5,04	**<0,0001**	10,39	4,49-24	**<0,0001**
Typ protezy (rozprężana na balonie vs samorozprężalna)	2,54	1,13-5,73	**0,02**	10,77	2,61-44,47	**<0,0001**

Współczynniki regresji i ich istotność statystyczna dla zależności opisującej związek pomiędzy koniecznością wszczepienia pacjentom stymulatora a wybranymi zmiennymi objaśniającymi.

Regresja logistyczna wieloczynnikowa

Dependent Y	Wszczepienie_stymulatora__1_nie__2_tak_

Method	Stepwise
Enter variable if P<	0,05
Remove variable if P>	0,1

Sample size	250
Positive cases [a]	39 (15,60%)
Negative cases [b]	211 (84,40%)

[a] Wszczepienie_stymulatora__1_nie__2_tak_ = 2
[b] Wszczepienie_stymulatora__1_nie__2_tak_ = 1

Overall Model Fit

Null model -2 Log Likelihood	216,489
Full model -2 Log Likelihood	100,951
Chi-squared	115,538
DF	5
Significance level	P < 0,0001
Cox & Snell R^2	0,3701
Nagelkerke R^2	0,6388

Coefficients and Standard Errors

Variable	Coefficient	Std. Error	Wald	P
Amiodaron__1_nie__2_tak_=2	2,87422	1,00375	8,1996	0,0042
Czas_trwania_QRS__ms_	0,054110	0,011172	23,4594	<0,0001
Głębokość_osadzenia_protezy_mm_	2,34039	0,42740	29,9857	<0,0001
RBBB__1_nie__2_tak_=2	1,60694	0,75355	4,5476	0,0330
Rodzaj_protezy__1_rozprężalna_na_balonie___2__samorozprężalna_=2	2,37708	0,72345	10,7963	0,0010
Constant	-25,92175	4,09111	40,1464	<0,0001

Variables not included in the model
Czas_trwania_PQ_ms_
LBBB__1_nie__2_tak_=2
Nitraty__1_nie__2_tak_=2
Wiek

Odds Ratios and 95% Confidence Intervals

Variable	Odds ratio	95% CI
Amiodaron__1_nie__2_tak_=2	17,7117	2,4766 to 126,6683
Czas_trwania_QRS__ms_	1,0556	1,0327 to 1,0790
Głębokość_osadzenia_protezy_mm_	10,3852	4,4938 to 24,0007
RBBB__1_nie__2_tak_=2	4,9875	1,1388 to 21,8433
Rodzaj_protezy__1_rozprężalna_na_balonie___2__samorozprężalna_=2	10,7734	2,6094 to 44,4799

Contingency table for Hosmer & Lemeshow test [Hide]

Group	Y=0		Y=1		Total
	Observed	Expected	Observed	Expected	
1	25	24,988	0	0,0116	25
2	25	24,965	0	0,0354	25
3	28	27,852	0	0,148	28
4	28	27,738	0	0,262	28
5	25	25,581	1	0,419	26
6	23	22,104	0	0,896	23
7	23	23,117	2	1,883	25
8	19	19,598	6	5,402	25
9	11	12,832	14	12,168	25
10	4	2,224	16	17,776	20

Classification table (cut-off value p=0,5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	204	7	96,68%
Y = 1	14	25	64,10%
Percent of cases correctly classified			91,60%

ROC curve analysis

Area under the ROC curve (AUC)	0,947
Standard Error	0,0158
95% Confidence interval	0,912 to 0,971

Regresja logistyczna wieloczynnikowa

	Analiza jednoczynnikowa			Analiza Wieloczynnikowa		
Parametr	**Iloraz szans**	**95%CI**	**P**	**Iloraz szans**	**95%CI**	**P**
Wiek	1,05	1-1,11	**0,05**			
Nitraty	4,74	1,56-14,49	**0,006**			
Amiodaron	3,85	1,19-12,39	**0,02**	17,71	2,48-126,67	**0,004**
LBBB	5,29	2,58-10,06	**<0,0001**			
RBBB	8,5	3,64-19,86	**<0,0001**	4,99	1,14-21,84	**0,03**
QRS (ms)	1,05	1,04-1,07	**<0,0001**	1,06	1,03-1,07	**<0,0001**
PQ (ms)	1,02	1,01-1,03	**0,004**			
Głębokość protezy w LVOT (mm)	3,3	2,16-5,04	**<0,0001**	10,39	4,49-24	**<0,0001**
Typ protezy (rozprężana na balonie vs samorozprężalna)	2,54	1,13-5,73	**0,02**	10,77	2,61-44,47	**<0,0001**

Analiza krzywej ROC (Receiver operating characteristic)

Contingency table for Hosmer & Lemeshow test [Hide]

Group	Y=0		Y=1		Total
	Observed	Expected	Observed	Expected	
1	25	24,988	0	0,0116	25
2	25	24,965	0	0,0354	25
3	28	27,852	0	0,148	28
4	28	27,738	0	0,262	28
5	25	25,581	1	0,419	26
6	23	22,104	0	0,896	23
7	23	23,117	2	1,883	25
8	19	19,598	6	5,402	25
9	11	12,832	14	12,168	25
10	4	2,224	16	17,776	20

Classification table (cut-off value p=0,5)

Actual group	Predicted group		Percent correct
	0	1	
Y = 0	204	7	96,68%
Y = 1	14	25	64,10%
Percent of cases correctly classified			91,60%

ROC curve analysis

Area under the ROC curve (AUC)	0,947
Standard Error	0,0158
95% Confidence interval	0,912 to 0,971

Regresja logistyczna wieloczynnikowa

Regresja logistyczna wieloczynnikowa

Variable	Głębokość_osadzenia_protezy_mm_
Classification variable	Wszczepienie_stymulatora__1_nie__2_tak_

Sample size	285
Positive group [a]	43 (15,09%)
Negative group [b]	242 (84,91%)

[a] Wszczepienie_stymulatora__1_nie__2_tak_ = 2
[b] Wszczepienie_stymulatora__1_nie__2_tak_ = 1

Disease prevalence (%)	15,1

Area under the ROC curve (AUC)

Area under the ROC curve (AUC)	0,813
Standard Error [a]	0,0281
95% Confidence interval [b]	0,762 to 0,856
95% Bootstrap CI [c]	0,755 to 0,863
z statistic	11,133
Significance level P (Area=0.5)	<0,0001

[a] Hanley & McNeil, 1982
[b] Binomial exact
[c] BC_a bootstrap confidence interval (1000 iterations; random number seed: 978).

Youden index

Youden index J	0,5568
Associated criterion	>6
Sensitivity	95,35
Specificity	14,46

Optimal criterion

Optimal criterion [a]	>6,5
Sensitivity	81,40
Specificity	14,46

[a] Taking into account disease prevalence (15,1%) and estimated costs:
cost False Positive: 0,5; cost False Negative: 1,5
cost True Positive: 0; cost True Negative: 0

Summary Table

Estimated specificity at fixed sensitivity			
Sensitivity	Specificity	95% CI [a]	Criterion
80,00	14,46	0,20 to 0,75	>6,515
90,00	14,46	0,20 to 0,75	>6,191666667
95,00	14,46	0,20 to 0,75	>6,0125
97,50	14,46	0,20 to 0,75	>5,76875
99,00	14,46	0,20 to 0,75	>5,6075
Estimated sensitivity at fixed specificity			
Specificity	Sensitivity	95% CI [a]	Criterion
80,00	59,69	44,98 to 76,20	>6,733333333
90,00	31,74	13,90 to 49,57	>7,1125
95,00	12,72	4,52 to 27,89	>7,884285714
97,50	6,69	0,00 to 0,00	>7,970714286
99,00	1,65	0,00 to 0,00	>8,645

[a] BC_a bootstrap confidence interval (1000 iterations; random number seed: 978).

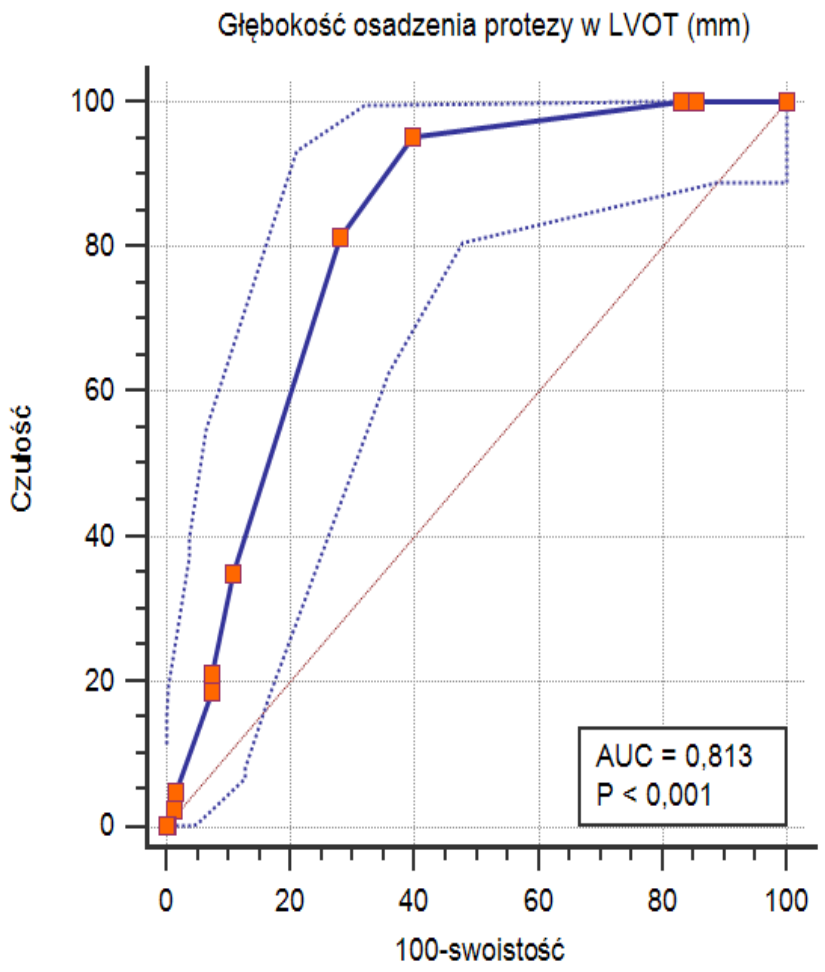
Criterion values and coordinates of the ROC curve [Hide]

Criterion	Sensitivity	95% CI	Specificity	95% CI	+LR	95% CI	-LR	95% CI	+PV	95% CI	-PV	95% CI	Cost
≥5	100,00	91,8 - 100,0	0,00	0,0 - 1,5	1,00	1,0 - 1,0			15,1	15,1 - 15,1			0,425
>5	100,00	91,8 - 100,0	14,46	10,3 - 19,5	1,17	1,1 - 1,2	0,00		17,2	16,5 - 18,0	100,0		0,363
>5,5	100,00	91,8 - 100,0	16,94	12,4 - 22,3	1,20	1,1 - 1,3	0,00		17,6	16,8 - 18,5	100,0		0,353
>6	95,35	84,2 - 99,4	60,33	53,9 - 66,5	2,40	2,0 - 2,8	0,077	0,02 - 0,3	29,9	26,5 - 33,6	98,6	94,9 - 99,6	0,179
>6,5	81,40	66,6 - 91,6	71,90	65,8 - 77,5	2,90	2,3 - 3,7	0,26	0,1 - 0,5	34,0	28,7 - 39,7	95,6	92,1 - 97,6	0,161
>7	34,88	21,0 - 50,9	89,26	84,7 - 92,9	3,25	1,9 - 5,6	0,73	0,6 - 0,9	36,6	25,0 - 49,9	88,5	86,1 - 90,6	0,193
>7,5	20,93	10,0 - 36,0	92,56	88,5 - 95,5	2,81	1,4 - 5,8	0,85	0,7 - 1,0	33,3	19,4 - 51,0	86,8	84,9 - 88,5	0,211
>7,8	18,60	8,4 - 33,4	92,56	88,5 - 95,5	2,50	1,2 - 5,4	0,88	0,8 - 1,0	30,8	17,1 - 48,9	86,5	84,7 - 88,1	0,216
>8	4,65	0,6 - 15,8	98,35	95,8 - 99,5	2,81	0,5 - 14,9	0,97	0,9 - 1,0	33,3	8,6 - 72,6	85,3	84,4 - 86,1	0,223
>8,5	2,33	0,06 - 12,3	98,76	96,4 - 99,7	1,88	0,2 - 17,6	0,99	0,9 - 1,0	25,0	3,4 - 75,8	85,1	84,4 - 85,7	0,226
>9	0,00	0,0 - 8,2	99,59	97,7 - 100,0	0,00		1,00	1,0 - 1,0	0,0		84,9	84,8 - 85,0	0,228
>9,5	0,00	0,0 - 8,2	100,00	98,5 - 100,0			1,00	1,0 - 1,0			84,9	84,9 - 84,9	0,226

Współczynniki regresji i ich istotność statystyczna dla zależności opisującej związek pomiędzy koniecznością wszczepienia pacjentom stymulatora a wybranymi zmiennymi objaśniającymi.

Krzywa ROC przedstawiająca wartość odcięcia dla głębokości osadzenia protezy zastawkowej w LVOT w predykcji konieczności implantacji stymulatora serca (wartość odcięcia - > 6,5 mm przy czułości – 81,40% i swoistości – 71,90% AUC – 0,813).

Regresja logistyczna wieloczynnikowa

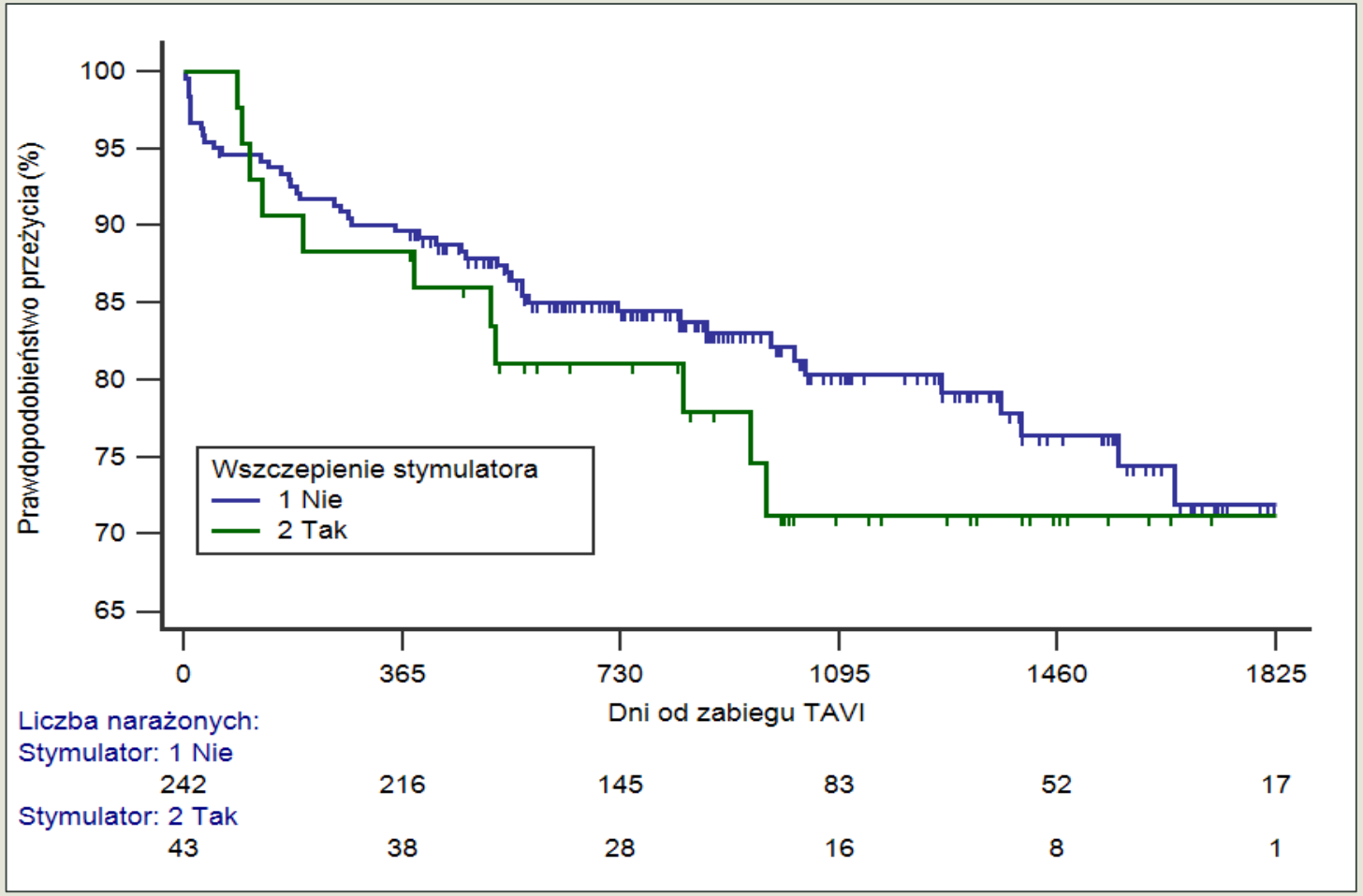


Współczynniki regresji i ich istotność statystyczna dla zależności opisującej związek pomiędzy koniecznością wszczepienia pacjentom stymulatora a wybranymi zmiennymi objaśniającymi.

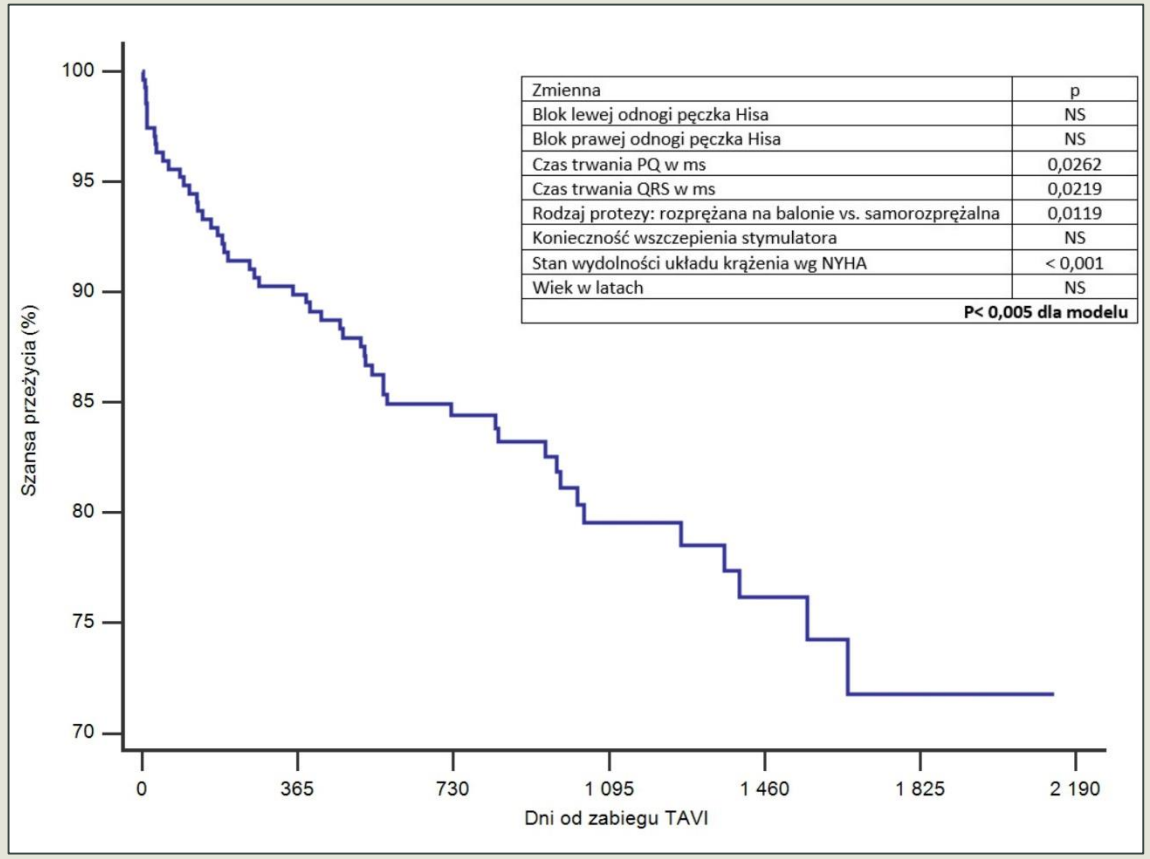
Krzywa ROC przedstawiająca wartość odcięcia dla głębokości osadzenia protezy zastawkowej w LVOT w predykcji konieczności implantacji stymulatora serca (wartość odcięcia - > 6,5 mm przy czułości – 81,40% i swoistości – 71,90% AUC – 0,813).

Analiza przeżycia
Estymację funkcji przeżycia - estymator Kaplana-Meiera

BP	CD
Wszczepienie_stymulatora__1_nie__2_tak_	Ile_dni_od_TAVI_do_zgonu_lub_do_końca_obserwacji
1	1020
1	62
2	468
1	861
1	710
2	1402
2	1325
1	836
1	1607
1	646
1	1117
2	34
1	98
1	460
1	176
1	426
1	785
1	29
1	274
2	952
1	433
1	1395
1	200
2	142
1	668
2	887
1	570
1	1117
1	998
1	646
1	385
1	



Analiza przeżycia - **Modele proporcjonalnego hazardu**
Model Coxa



Zmienna	p
Blok lewej odnogi pęczka Hisa	NS
Blok prawej odnogi pęczka Hisa	NS
Czas trwania PQ w ms	0,0262
Czas trwania QRS w ms	0,0219
Rodzaj protezy: rozprężana na balonie vs. samorozprężalna	0,0119
Konieczność wszczepienia stymulatora	NS
Stan wydolności układu krążenia wg NYHA	< 0,001
Wiek w latach	NS
	P< 0,005 dla modelu

Regression

Description

Regression is a method used to describe the relationship between two variables and to predict one variable from another (if you know one variable, then how well can you predict a second variable?).

Whereas for correlation the two variables need to have a Normal distribution, this is not a requirement for regression analysis. The variable X does not need to be a random sample with a Normal distribution (the values for X can be chosen by the experimentor). However, the variability of Y should be the same at each level of X.

The program will calculate the following statistics:

- Sample size: the number of (selected) data pairs
- Coefficient of determination: this is the proportion of the variation in the dependent variable explained by the regression model. It can range from 0 to 1 and is a measure of the goodness of fit of the model.
- Residual standard deviation: the standard deviation of the residuals
- The regression equation: the selected equation with the calculated coefficients.
- Analysis of variance table, with F-ratio and P value.

Required input

Variables

- **Variable Y** and **Variable X**: select the dependent and independent variables Y and X.

Regression equation

By default the option **Include constant in equation** is selected. This is the recommended option that will result in ordinary least-squares regression. When you need regression through the origin (no constant a in the equation), you can uncheck this option (an example of when this is appropriate is given in Eisenhauer, 2003).

MedCalc offers a choice of 5 different regression equations (x represents the independent variable and y the dependent variable):

y	=	a + b x	straight line
y	=	a + b log(x)	logarithmic curve
log(y)	=	a + b x	exponential curve
log(y)	=	a + b log(x)	geometric curve
y	=	a + b x + c x^2	quadratic regression (parabola)

Results

The results window for Regression displays:

Sample size: the number of (selected) data pairs

Coefficient of determination R^2: this is the proportion of the variation in the dependent variable explained by the regression model. It can range from 0 to 1 and is a measure of the goodness of fit of the model.

Residual standard deviation: the standard deviation of the residuals (residuals = differences between observed and predicted values)

The regression equation: the selected equation with the calculated values for intercept *a* and slope *b* (and for a parabola a third coefficient *c*). E.g.

$$y = a + b\,x$$

The standard errors are given for the intercept *a* and the slope *b*, followed by the t-value and the P-value for the hypothesis that these coefficients are equal to 0. If the P-values are low (e.g. less than 0.05), then you can conclude that the coefficients are different from 0.

Note that when you use the regression equation for prediction, you may only apply it to values in the range of the actual observations. E.g. when you have calculated the regression equation for height and weight for school children, this equation cannot be applied to adults.

Analysis of variance table, with F-ratio and P value.

The analysis of variance table divides the total variation in the dependent variable into two components, one which can be attributed to the regression model (labeled Regression) and one which cannot (labelled Residual). If the significance level for the F-test is small (less than 0.05), then the hypothesis that there is no (linear) relationship can be rejected.

Literature

- Armitage P, Berry G, Matthews JNS (2002) Statistical methods in medical research. 4th ed. Blackwell Science.
- Bland M (2000) An introduction to medical statistics, 3rd ed. Oxford: Oxford University Press.
- Eisenhauer JG (2003) Regression through the origin. Teaching Statistics 25:76-80.
- Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) Applied linear statistical models. 4th ed. Boston: McGraw-Hill.

Analiza regresji

Analiza regresji