



Technical University of Lodz

Institute of Electronics

Biostatystyka

Statystyka biomedyczna

Wprowadzenie

Piotr M. Szczypiński & Artur Klepaczko

Dane kontaktowe

Piotr M. Szczypiński, dr hab. inż.

Instytut Elektroniki, Politechnika Łódzka

Wólczańska 211/215, 90-924 Łódź

Pokój 217A

Konsultacje:

Środy 10:15 – 11:00

tel. +48 42 6312642

<http://www.eletel.p.lodz.pl/pms>

piotr.szczypinski@p.lodz.pl



Dane kontaktowe

Artur Klepaczko, dr inż.

Instytut Elektroniki, Politechnika Łódzka

Wólczańska 211/215, 90-924 Łódź

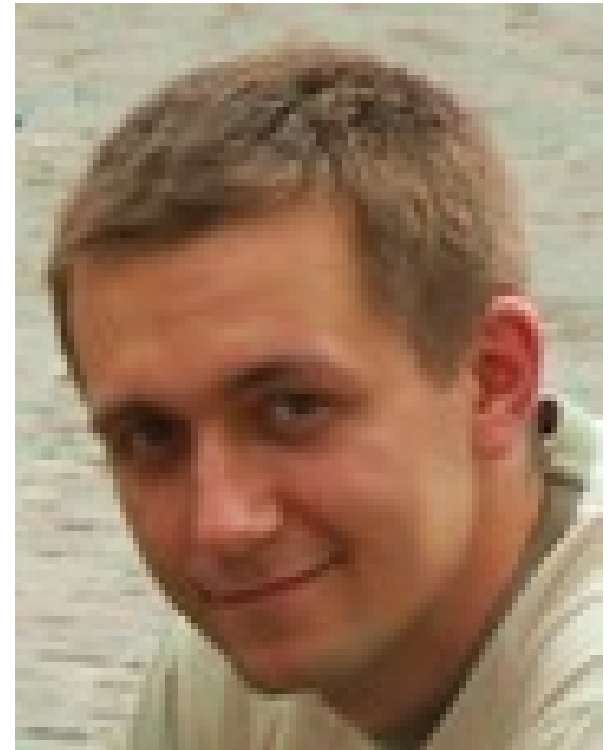
Pokój 217A

Konsultacji:

tel. +48 42 6312642

<http://eletel.p.lodz.pl/aklepaczko>

artur.klepaczko@p.lodz.pl



Cel przedmiotu

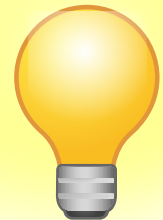
Poznanie podstawowych metod analizy statystycznej danych i zjawisk.

Nabywanie umiejętności analizy danych z zastosowaniem metod statystyki.



Efekty kształcenia

- 1) Po zaliczeniu przedmiotu student potrafi rozpoznawać różne rozkłady zmiennych losowych i estymować ich parametry
- 2) Wyznaczać ilościowe zależności pomiędzy zmiennymi losowymi
- 3) Stosować testy istotności i wyznaczać przedziały ufności
- 4) Prezentować graficznie wyniki analiz statystycznych i je interpretować
- 5) Testować hipotezy statystyczne
- 6) Stosować metody selekcji i ekstrakcji cech
- 7) Klasyfikować dane
- 8) Stosować ilościowe miary oceny klasyfikacji



Zakres wykładu

Czym jest biostatystyka.

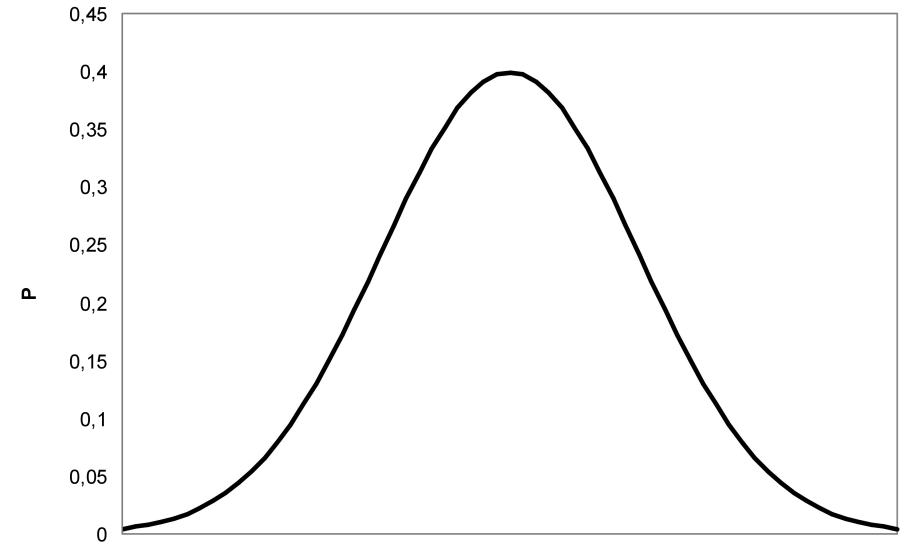
Definicja prawdopodobieństwa. Zależność zdarzeń i prawdopodobieństwa warunkowe. Rozkłady zmiennych losowych ciągłych i dyskretnych; dystrybuanta i funkcja gęstości prawdopodobieństwa. Rozkład normalny.

Wprowadzenie do metod Monte-Carlo.

Miary rozkładów: momenty, moda, mediana, percentyle, kurtoza, itp. Rozkład zmiennych losowych dwuwymiarowych. Zależność zmiennych losowych, współczynnik korelacji i macierz korelacji.

Graficzna prezentacja wyników: wykresy pudełkowe, siatki centylowe, krzywa ROC.

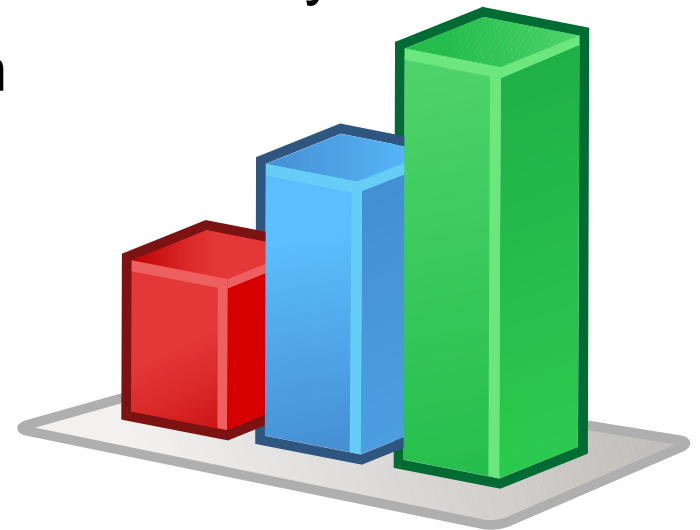
Testowanie hipotez statystycznych, testy statystycznej istotności i niezależności. Podstawy analizy regresji w zastosowaniach biomedycznych. Klasyfikacja danych i analiza skupień. Miary oceny klasyfikacji. Selekcja i ekstrakcja cech znaczących.



Zakres laboratorium

Rozwiązywanie praktycznych problemów z zastosowaniem narzędzi statystyki matematycznej w języku Python:

1. Graficzna prezentacja danych statystycznych
2. Zastosowania rozkładów zmiennych losowych
3. Badanie zgodności rozkładów z rozkładem normalnym
4. Badanie zależności zmiennych losowych
5. Analiza regresji
6. Testowanie hipotez statystycznych
7. Zastosowania symulacji Monte-Carlo
8. Klasyfikacja danych i analiza skupień.
9. Projekt końcowy

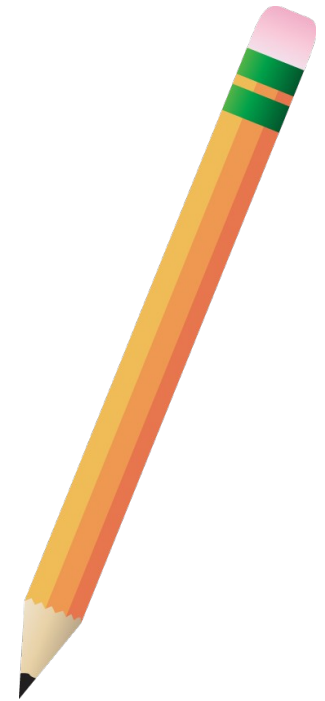


Podstawa zaliczenia

Efekty kształcenia są weryfikowane na podstawie testu pisemnego (w zakresie wiedzy) oraz na podstawie wykonywanych ćwiczeń laboratoryjnych i zadań projektowych (w zakresie umiejętności)

Wykład: pisemne zaliczenie dotyczące zagadnień przedstawianych w trakcie wykładu.

Laboratorium: ocena wykonanych ćwiczeń laboratoryjnych raporty z projektów. Ocena końcowa stanowi ważoną średnią arytmetyczną z ocen uzyskanych dla obydwu form realizacji przedmiotu.



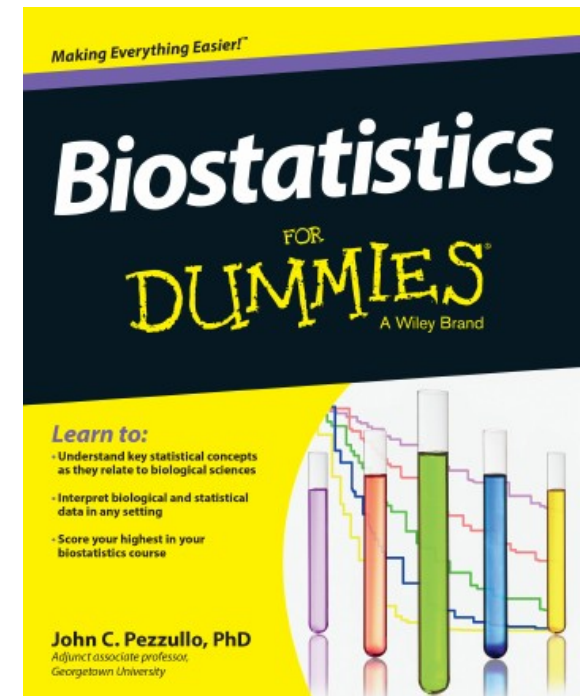
Literatura uzupełniająca

A. Papoulis, U. Pillai - Probability, Random Variables and Stochastic Processes, McGraw Hill, 2002

S. Brandt - Analiza danych, Wydawnictwo Naukowe PWN, Warszawa 1999

R. Johnson, D. Wichern - Applied Multivariate Statistical Analysis, Pearson Education Limited, 2014

John C. Pezzullo - Biostatistics For Dummies, John Wiley & Sons, 2013



Materiały w sieci

<https://medschool.vanderbilt.edu/biostatistics/content/what-biostatistics>

<http://sebastianraschka.com/Articles/>

<http://blog.yhathq.com/posts/logistic-regression-and-python.html>

http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

<http://glowingpython.blogspot.com/2012/03/linear-regression-with-numpy.html>

http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

<http://aimotion.blogspot.com/2011/10/machine-learning-with-python-linear.html>

<http://www.turingfinance.com/regression-analysis-using-python-statsmodels-and-quandl/>

<http://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.linregress.html>

<http://glowingpython.blogspot.com/2012/03/linear-regression-with-numpy.html>

Statystyka

Statystyka, nauka zajmująca się metodami badania przedmiotów lub zjawisk masowych w przestrzeni lub w czasie i ich ilościową lub jakościową analizą z punktu widzenia nauki do której należą

Encyklopedia Powszechna PWN

Historia

Słowa „statystyka”, „statystyk”, „statystyczny” w ich obecnym znaczeniu używane są od ok. stu lat.

Słowo „statystyka” pochodzi od łacińskiego status, oznaczającym stan państwa. Początki gałęzi wiedzy, zwanej statystyką, sięgają XVI wieku, kiedy we Włoszech zaczęto opisywać stan państwa (stato) za pomocą liczb.

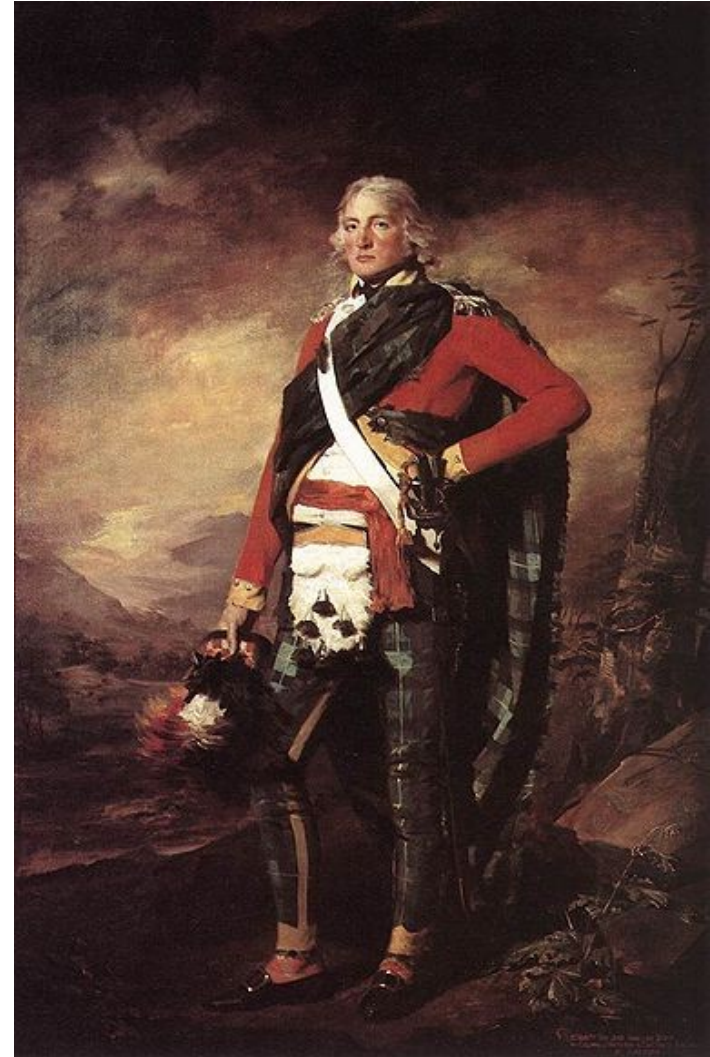
Statystyka rozumiana jako metoda liczbowego opisu stanu państwa, rozwinęła się jednak dopiero w XVII i XVIII wieku, głównie w Niemczech i Anglii.

Małgorzata Radzukiewicz, Wykład Statystyka

Historia

Sir John Sinclair (1754-1835)

angielski ekonomista i agronom, założyciel i pierwszy prezes angielskiego urzędu do spraw rolnictwa, wydawca i organizator pierwszego rachunku statystycznego Szkocji (1791-99). W liście do duchownych Kościoła Szkocji (1790) wyjaśnia, że w Niemczech „...badania statystyczne (...) są prowadzone na bardzo szeroka skalę” i dodaje iż badania statystyczne dotyczą „ludności, warunków politycznych, produkcji kraju i innych spraw państwa.” W swojej pionierskiej 21 tomowej pracy „Statistical Accounts of Scotland” jako pierwszy użył słowa statystyka w języku angielskim



Historia

Z biegiem czasu metody opracowane pierwotnie dla potrzeb analizy stanu państwa wykorzystywano do analizy innego rodzaju danych. W ten sposób powstała dziedzina wiedzy zwana **statystyka opisową**.



W okresie, gdy powstawały metody statystyki opisowej, rozwijana była równoległe metodologia **rachunku prawdopodobieństwa**.



Połączenie tych dwóch dziedzin wiedzy zaowocowało powstaniem **statystyki matematycznej**, zwanej też **teorią wnioskowania statystycznego**.

Dziedziny

statystyka

```
graph TD; statystyka --> matematyczna; statystyka --> opisowa; matematyczna --> teoria_estymacji[teoria estymacji]; matematyczna --> teoria_weryfikacji[teorię weryfikacji hipotez statystycznych];
```

matematyczna

metody poznania własności rozkładu jednej lub kilku cech w zbiorowości statystycznej (populacji generalnej) na podstawie zbadania części tej zbiorowości (zwanej próbą lub próbką), wybranej w sposób losowy.

opisowa

metody gromadzenia danych statystycznych, (zbiorowości statystyczne), prezentacji danych statystycznych (w sposób tabelaryczny i graficzny), sumarycznego opisu danych statystycznych (za pomocą parametrów statystycznych)

teoria estymacji

teorię weryfikacji
hipotez statystycznych

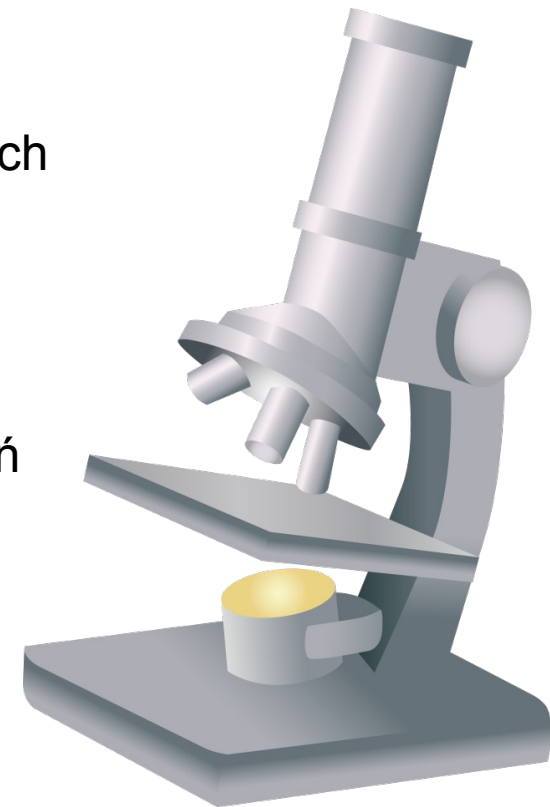
Biostatystyka

Biostatystyka – nauka z pogranicza biologii, medycyny i statystyki, adaptacja metod statystycznych na potrzeby prac badawczych w dziedzinie biologii, związanych przede wszystkim z medycyną, genetyką, fizjologią, antropologią, ekologią i rolnictwem.

Przykładowymi zastosowaniami biostatystyki są badania obserwacyjne, badania promujące leki oraz usługi analizy danych dla badań klinicznych, obserwacyjnych oraz badań biorównoważności.

Prace badawcze prowadzone w ramach biostatystyki obejmują: projektowanie eksperymentów biologicznych, zbieranie, agregowanie i analizowanie danych pochodzących z tych badań oraz interpretowanie wyników i formułowanie wniosków.

Biostatystyka [online]. Wikipedia : wolna encyklopedia





DEPARTMENT OF BIostatISTICS

Home

What is Biostatistics?

About the Department

Our People

Biostatistics Graduate
Program

Biostatistics
Collaboration Center
(BCC)

Biostatistics Wiki

Department of Biostatistics

Vanderbilt School of Medicine
2525 West End, Ste. 11000
Nashville, TN 37203
615-322-2001 (phone)
615-343-4924 (fax)

Cancer Biostatistics Division

571 Preston Research Bldg
Nashville, TN 37232-6848
615-936-2572 (phone)
615-936-2602 (fax)

What is Biostatistics?

Biostatistics is the branch of statistics responsible for the proper interpretation of scientific data generated in the biology, public health and other health sciences (i.e., the biomedical sciences). In these sciences, subjects (patients, mice, cells, etc.) exhibit considerable variation in their response to stimuli. This variation may be due to different treatments or it may be due to chance, measurement error, or other characteristics of the individual subjects. Biostatistics is particularly concerned with disentangling these different sources of variation. It seeks to distinguish between correlation and causation, and to make valid inferences from known samples about the populations from which they were drawn. (For example, do the results of treating patients with two therapies justify the conclusion that one treatment is better than the other?) Biostatisticians regularly engage in interdisciplinary collaborations and advance their discipline through pursuit of methodological development of statistical methods for biomedical applications. Biostatistics is a broad discipline encompassing the application of statistical theory to real-world problems, the practice of designing and conducting biomedical experiments and clinical trials (experiments with human subjects), the study of related computational algorithms and display of data, and the development of mathematical statistical theory. Biostatistics is integral to the advance of knowledge in biology, health policy, clinical medicine, public health policy, health economics, proteomics, genomics, and other disciplines. At Vanderbilt, biostatisticians facilitate biomedical research by providing methodological expertise and by closely collaborating with scientists and physician researchers.

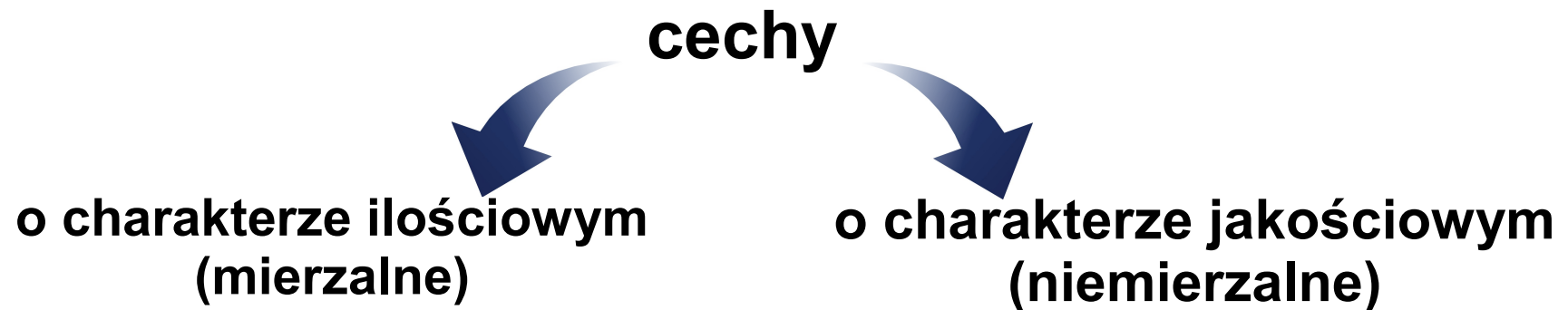
Biostatisticians are specialists in the evaluation of data as scientific evidence. They understand the generic construct of data and they provide the mathematical framework that transcends the scientific context to generalize the findings. Their expertise includes the design and conduct of experiments, the mode and manner in which data are collected, the analysis of data, and the interpretation of results. Meaningful generalization of experimental results requires the application of an appropriate mathematical framework for the scientific context. The validity of research results depends on this application and the reproducibility of the experimental methods. Biostatisticians use mathematics to enhance science and bridge the gap between theory and practice.

<https://medschool.vanderbilt.edu/biostatistics/content/what-biostatistics>

Podstawowe pojęcia

Przedmiotem statystyki jest **populacja generalna** zwana też **zbiorowością statystyczną** (zbiór jednostek, elementów zbiorowości)

Cecha statystyczna to właściwość pozwalająca odróżnić jednostki zbiorowości



Metody statystyki matematycznej

Analiza wariancji

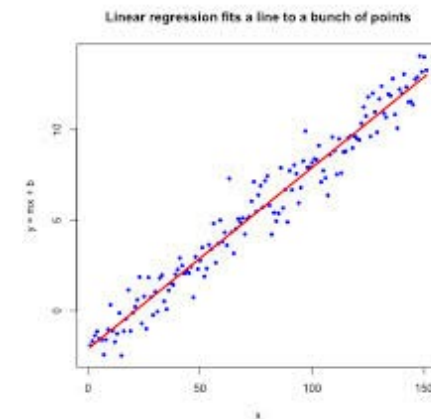
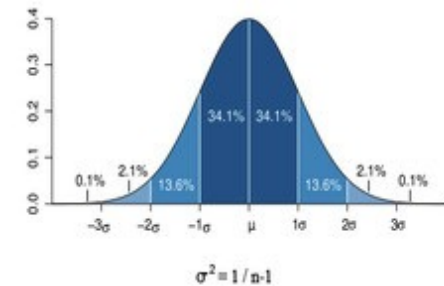
to technika postępowania statystycznego stosowana przy badaniu systematycznego wpływu danych czynników na przypadkowe wyniki. Analiza wariancji dostarcza kryteriów pozwalających rozstrzygnąć czy dane uzyskane z obserwacji przeprowadzonego eksperymentu można uznać za przypadkowe, czy należy je uznać za odchylenia systematyczne.

Analiza współzależności

to jednoczesne badanie populacji ze względu na dwie lub więcej cech (analiza regresji i analiza korelacji)

Metoda reprezentacyjna

wyznacza charakterystyki badanej zbiorowości statystycznej bez statystycznego opracowywania całej zbiorowości, lecz na podstawie danych z próby (z reprezentacji).



Przykład - analiza

Przykład: zbiorowość ziaren jęczmienia.

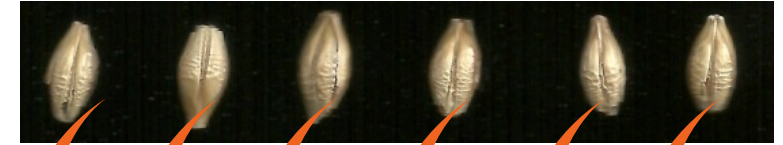
Element zbiorowości to pojedynczy ziarniak.

Ziarniak ma cechy niemierzalne (jakościowe), np. przynależność odmianowa, rodzaj uszkodzenia.

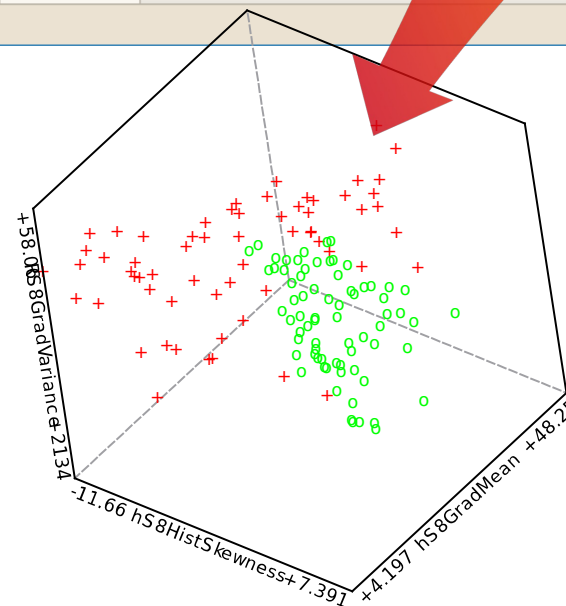
Ziarniak ma cechy mierzalne (ilościowe) takie jak kolor, proporcje kształtu, stopień pofałdowania.

Wektor cech mierzalnych dla każdego elementu zbiorowości można przedstawić jako punkt w przestrzeni cech.

Można badać zmienność cech mierzalnych, zależności między cechami mierzalnymi a także zależności między mierzalnymi i niemierzalnymi.



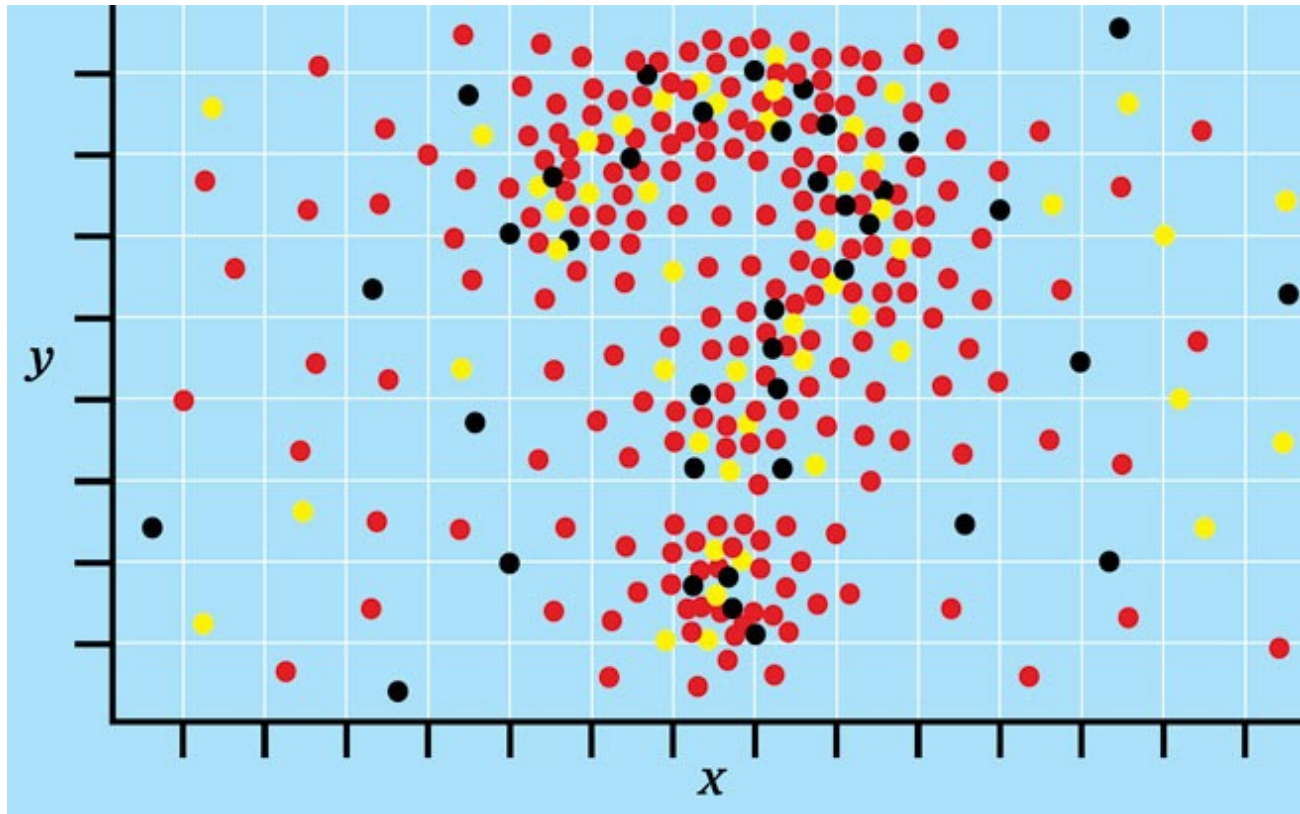
	✓ Normal	✓ Normal	✓ Normal	✓ Fusarium	✓ Fusarium	✓ Fusarium
✓ YS8Gab4V4Mag	20.8696	22.1299	18.4636	13.7746	30.26	20.4211
✓ YS8Gab4V8Mag	3125.2	3134.92	3136.7	3138.09	3130.18	3132.29
✓ YS8Gab4N4Mag	8.16444	8.80729	7.85408	7.74082	8.20131	9.28458
✓ YS8Gab4N8Mag	10.1505	11.8594	10.4445	9.88661	10.6138	11.8443
✓ YS8Gab4Z4Mag	8.69674	12.7013	8.71384	8.60365	10.3237	10.5996
✓ YS8Gab4Z8Mag	14.5428	19.6433	13.3026	13.176	16.2732	15.583
✓ YS8Gab80Mag	6148.67	6176.26	6181.54	6184.8	6155.92	6162.84
✓ YS8Gab8H2Mag	724.811	705.812	737.954	698.015	728.078	711.071
✓ YS8Gab8H4Mag	97.7045	71.8914	95.7805	61.6328	74.3683	92.8163
✓ YS8Gab8H8Mag	32.1716	30.2525	33.8331	18.7857	34.3166	38.58



Przestrzeń trzech wybranych cech

○ ziarniaki zdrowe
+ ziarniaki z fuzarium

Wnioskowanie



http://www.wired.com/2010/04/st_thompson_statistics/