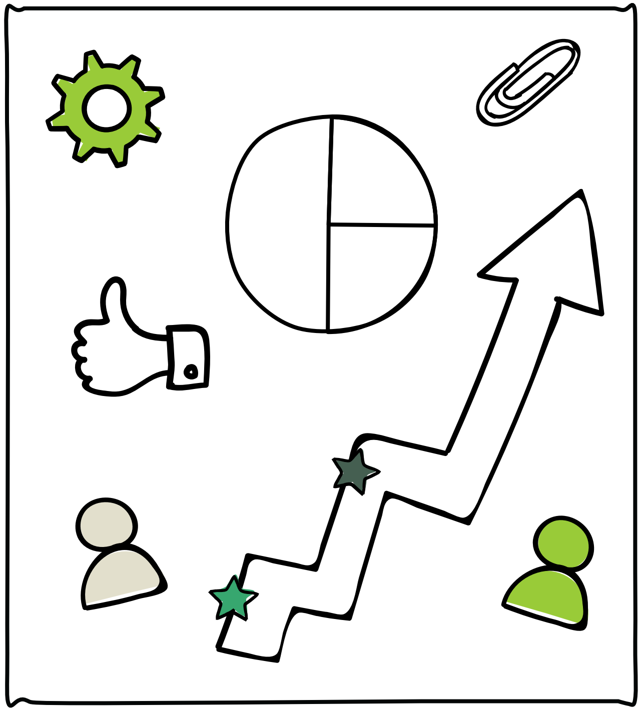


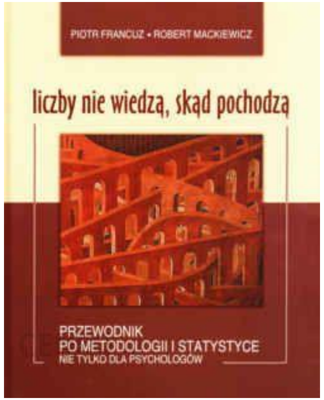
Statystyka

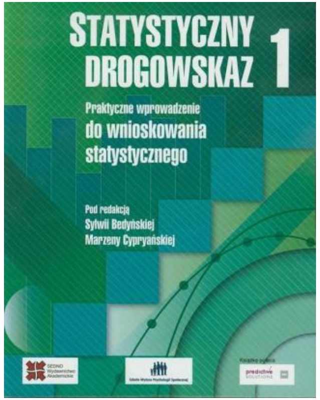
Natalia Józefacka

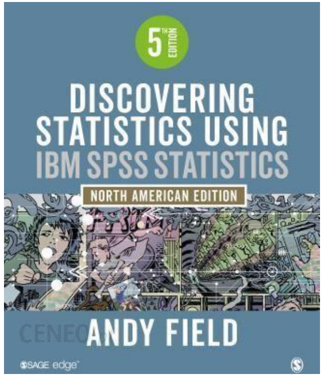
Plan wykładu

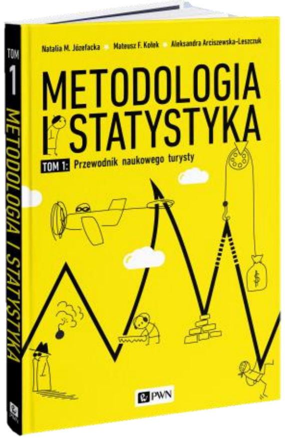
1. Prawdopodobieństwo (1h)
2. Podstawy wnioskowania statystycznego (2h)
    1. Próba vs populacja
    2. Szacowanie
    3. Rozkład
    4. Testowanie hipotez
3. Statystyki opisowe (2h)
4. Testowanie hipotez dotyczące średnich (2h)
5. Testowanie hipotez dotyczących związku (2h)

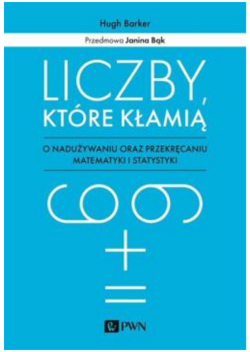
Literatura









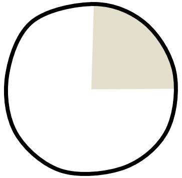


https://statystykawpsychologii.blogspot.com/

Statystyka – po co mi to?

- https://www.youtube.com/watch?v=ogeGJS0GEF4

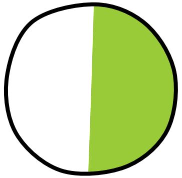
Wprowadzenie do wprowadzenia



25%

Saturn

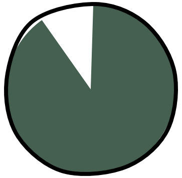
Yes, this is the ringed one



50%

Jupiter

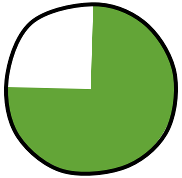
Jupiter is the biggest planet



95%

Mars

Despite being red, Mars is cold



75%

Venus

Venus has a beautiful name

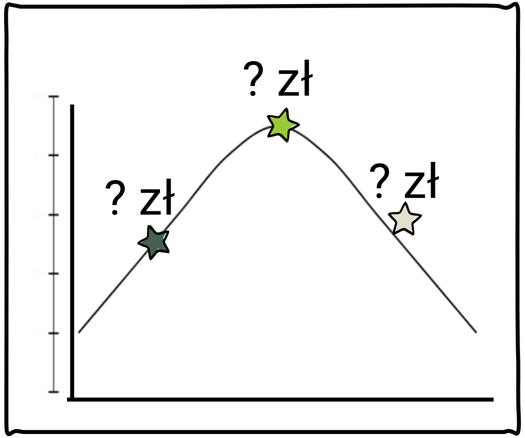
Podstawy wnioskowania statystycznego

- Próba vs populacja
- Szacowanie
- Rozkład
- Testowanie hipotez

Wprowadzenie

- To najtrudniejsza część, ale umożliwi Ci zrozumienie wszystkich kolejnych
- Pozwala na ocenę prawdopodobieństwa wystąpienia zjawisk, w sposób obiektywny
- Nie jest to intuicyjny sposób myślenia! Nasze heurystyki nie mają tu zastosowania!

Case study



702 oferty pracy

Populacja a próba

- Populacja – to całkowity zbiór interesujących nas elementów np. ogłoszenia o pracę, osoby mieszkające na terenie Polski, uczniowie klasy 5b itp.

702 oferty pracy

Czy ktoś może podać przykład innych populacji?

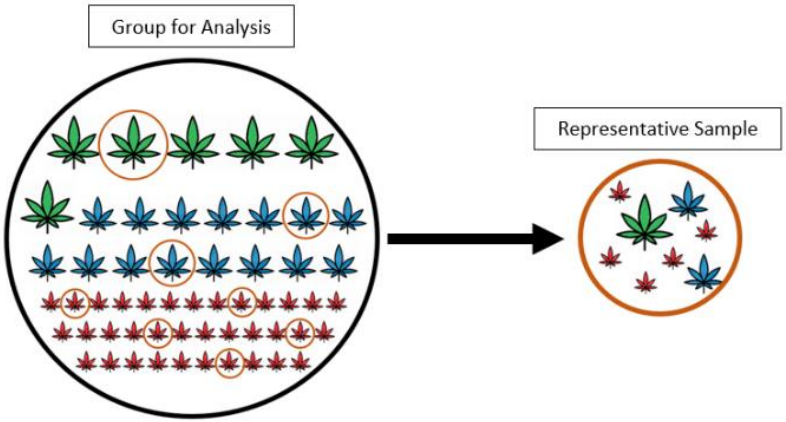
Czy jest możliwość zbadania całej populacji?

Próba

- Próba badawcza jest dowolną (choć najlepiej wystarczająco dużą) liczbą obserwacji, które pobiera się z interesującego badacza zbioru wszystkich elementów, po to, aby móc na ich podstawie wnioskować o całej populacji.
- Pobierają próbę *n* elementów z populacji – które będą ją reprezentować
  - Czyli będą **reprezentatywne!**

Reprezentatywność próby

- **Reprezentatywność** próby oznacza zgodność jej charakterystyki (czyli cech) z charakterystyką populacji, z której została pobrana.
- W tym celu należy wylosować próbę z populacji



Losowy dobór próby

Próba losowa - to próba pobrana w taki sposób, aby każda jednostka będąca częścią całej populacji, z której tę próbę się pobiera, miała jednakowe prawdopodobieństwo znalezienia się w niej.

Przed losowaniem:

1. Określ populację badawczą
2. Określ operat losowania – wg jakiego schematu będziesz losował, numery ogłoszeń o pracę, numer PESEL,
3. Ustal liczebność próby
4. Wybierz adekwatną metodę

Sposoby losowania

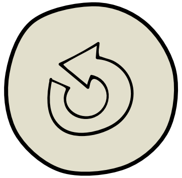
Losowanie proste

- Bezzwrotne
- Zwrotne

Losowanie systematyczne

Losowanie warstwowe

Sposoby losowania



Losowanie proste

- **Bezzwrotne**
- **Zwrotne**

- Wykorzystujesz do tego tablice liczb losowych lub generator liczb np. randomiser.org
- Wybierasz z jakiej populacji losujesz – 702 ogłoszenia o pracę
- ile chcesz wylosować obiektów – 30 ogłoszeń
  - Bezzwrotne: raz wylosowany obiekt wypada z puli
  - Zwrotne: każdy wylosowany obiekt wraca do puli i jest szansa wylosować go ponownie

Sposoby losowania

Losowanie systematyczne

- Wybierasz schemat losowania np.
    - losowanie co 5 ogłoszenia w bazie
    - Albo co 10
- Upewnij się że kolejność jest przypadkowa!

Sposoby losowania

Losowanie warstwowe

- Stosowane gdy populacja jest bardzo zróżnicowana
- Tobie zależy na dokładnych danych
- Losowanie na podstawie jakiegoś kryterium
  - Stanowiska: junior, regular, senior
- Wtedy z każdej warstwy losujesz określoną liczbę elementów, powinna być taka sama dla każdej warstwy (chyba że połączysz to z doborem kwotowym)

Jaki system losowania zastosowaliście w swoich badaniach?

Metody nielosowe

Dobór kwotowy

Metoda kuli śnieżnej

Dobór celowy

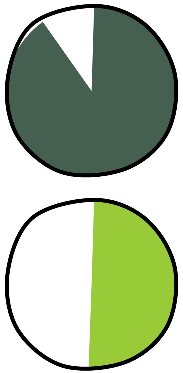
Dobór przypadkowy

Dobór przez eliminację

Dobór jednostek typowych

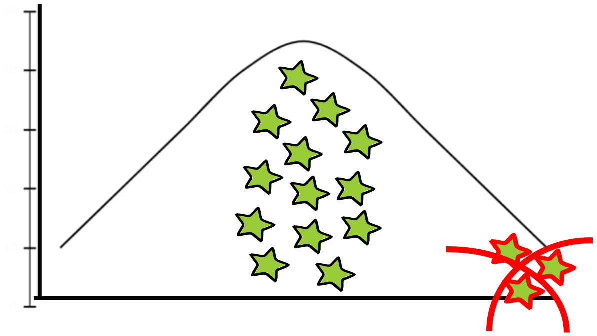
Dobór kwotowy

- Jeżeli znasz strukturę populacji, np. wśród **matematyków-statystyków** występuje 80% mężczyzn i 20% kobiet,
- a wśród **psychologów-statystyków** relacja jest 50/50.
- Chcąc dowiedzieć się, ile zarabiają poszczególni statystycy
  - Dobierasz próbkę z każdej populacji o wskazanym stosunku K i M



Dobór jednostek typowych

- wybiera się jednostki, które badacz uważa za typową reprezentację populacji, np. wybór ogłoszeń o zatrudnieniu na stanowiskach statystyka w przeciętnych przedsiębiorstwach.

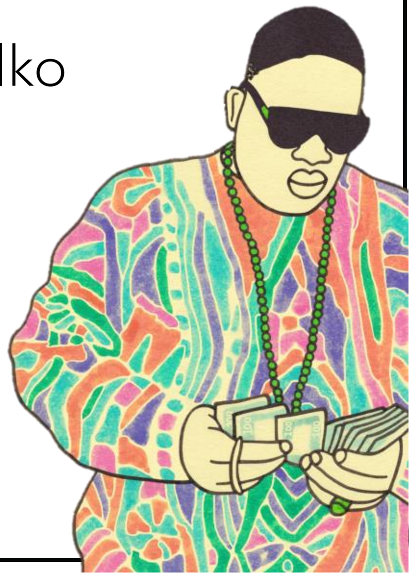


Dobór przez eliminację

Podobnie jak poprzednio, badacz eliminuje się z próby jednostki nietypowe

Metoda kuli śnieżnej

- Jeżeli populacja jest trudno dostępna
- Znajdujesz małą grupę należącą do populacji i prosisz o przekazanie informacji dalej
- Np. zarobki statystyków na czarnym rynku i znasz tylko dwóch pracujących dla mafii…



Dobór celowy

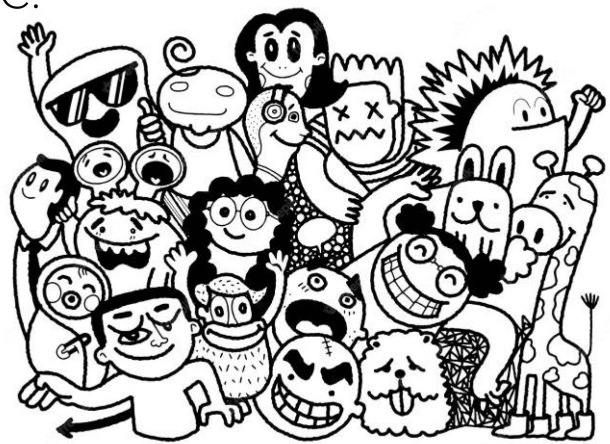
- Wybierasz próbkę zgodnie z założonymi celami np. ogłoszenia ale tylko dla Krakowa
- Dzieci ale tylko ze szkoły nr XX

Dobór przypadkowy

- Bierzesz kogo znajdziesz
- Metodologicznie wątpliwe!

ZAMIAST:

MOŻESZ MIEĆ:



Liczebność próby

- Reprezentatywność próby ma związek z precyzją wnioskowania
- Aby próba była reprezentatywna musi być też dostatecznie duża
- Celem NIE JEST dokładne określenie prawdziwej średniej zarobków, LECZ precyzyjne oszacowanie na podstawie jedynie części ogłoszeń o pracę
- Ważne abyś przejrzał 30*, a nie np. 5 ogłoszeń

Szacowanie czyli estymacja

- W statystyce niczego nie będziesz wiedział na 100%
- Wszystkie dane są obciążone błędem
- Dlatego **sceptyzm naukowy** z poprzedniego semestru dalej jest w cenie
- losowość zjawisk jest wpisana w definicję analiz statystycznych, dlatego właśnie statystykę tak wiele łączy z **teorią prawdopodobieństwa**

Prawdopodobieństwo

- To stosunek uzyskania pożądanego wyniku do ilości wszystkich możliwości
- Np. obstawiając szczęśliwy numerek w lotto
  obstawiasz jeden numer
  a możliwości jest 36
- Prawdopodobieństwo wynosi 1/36

Jeżeli kobieta jest w ciąży jakie jest prawdopodobieństwo że urodzi się dziewczynka?

Ilość możliwości 2 (chłopiec/dziewczynka)

Dziewczynka to 1 z 2 możliwości

Więc prawdopodobieństwo wynosi 1/2

Podejście częstościowe

Wróćmy do LOTTO i szczęśliwego numerka

- Prawdopodobieństwo wygrania wynosi 1/36 (ok 3%)
- częstość losowania poszczególnych liczb w kolejnych losowaniach ze zwracaniem zmienia się.
- powtarzając takie losowanie np. 36 razy, raczej nie wylosujesz każdej liczby z zakresu 1-36 (przeprowadzając nawet 1000 takich losowań nie sprawi to, że każda liczba zostanie wylosowana dokładnie tyle samo razy).
- Jednak! Różnice w liczbie wystąpień określonych liczb akurat w tego typu losowaniach nie są duże!

Podejście częstościowe

- Rzucając 3x monetą – może się zdarzyć, że:
  - 3x będzie orzeł,
  - 3xreszka
  - 2xorzeł 1x reszka

  mimo że za każdym razem prawdopodobieństwo wynosi ½
- ALE! Rzucając 100 razy – rozkład będzie zbliżał się do płaskiego, czyli 1/2

Eureka!

Nie zawsze musisz wiedzieć ile wynosi twoja populacja.

Ważne! Z każdą powtórzoną obserwacją jesteś bliżej odkrycia ile wynosi prawdopodobieństwo.

Dzięki temu możesz badać pewne wzorce rzeczywistości, w sytuacji gdy niewiele o nich wiemy.

Po co nam to?

Możesz sobie wyobrazić, że np. szansa istnienia takiej pracy w zawodzie statystyka, za którą pracodawca płaci 20 000 zł, nie jest już tak duża, jak ta, w której oferuje pensję 4000 zł, czyż nie?

To zróżnicowane prawdopodobieństwo przekłada się w końcu na to, jaką realnie pracę można znaleźć!

Wnioskowanie o zarobkach

Możesz przeprowadzić na dwa sposoby, wykorzystując:

- Estymację punktową – np. wyliczając średnią z ofert, otrzymasz wtedy 1 punkt
- Estymację przedziałową – otrzymasz wtedy zakres kwot tzw. widełki w których masz szansę zarabiać

Estymacja punktowa *n* = 5

Oferta	Proponowane zarobki
1	5000 zł
2	6000 zł
3	7000 zł
4	10000 zł
5	7000 zł
Średnia	**7000 zł**

- Wylosowano 5 ogłoszeń
- M = 7000
- Max. = 10000
- Min. = 5000

Precyzja szacowania jest niska

Estymacja punktowa *n* = 10 i *n* = 30 i prawdziwy

- n = 10
- M = 5200
- Min. = 3000
- Max. = 7000

PRAWDZIWY WYNIK
N = 702
M = 5720,80 zł

- n = 30
- M = 6000
- Min. = 3500
- Max. = 11000

Parametr

- $n = 10$
- $M = 5200$
- Min. = 3000
- Max. = 7000

- $n = 30$
- $M = 6000$
- Min. = 3500
- Max. = 11000

PRAWDZIWY WYNIK
$N = 702$
$M = 5720{,}80$ zł

Parametr: konkretna kwota która charakteryzuje daną skończoną populację*

Populacja skończona vs hipotetyczna

- n = 10
- M = 5200
- Min. = 3000
- Max. = 7000

- n = 30
- M = 6000
- Min. = 3500
- Max. = 11000

PRAWDZIWY WYNIK
N = 702
M = 5720,80 zł

Parametr: konkretna kwota która charakteryzuje daną skończoną populację*

***Populacja skończona** to taka populacja, która istnieje i obejmuje określoną liczbę jednostek; w większości przypadków badań wnioskujemy jednak z próby pobranej nie z populacji skończonej, ale **populacji hipotetycznej**. Z populacji skończonej można obliczyć prawdziwą wartość, a z hipotetycznej pozostają nieznane!

Estymator

- $n = 10$
- $M = 5200$
- Min. = 3000
- Max. = 7000

- $n = 30$
- $M = 6000$
- Min. = 3500
- Max. = 11000

Estymator: to wartość z próby, mająca odzwierciedlić wartość parametru.

PRAWDZIWY WYNIK
$N = 702$
M = 5720,80 zł

Parametr: konkretna kwota która charakteryzuje daną skończoną populację*

Błąd pomiaru

- Różnica między **szacowanym wynikiem**

$n = 5$
$M = 7000$

$n = 10$
$M = 5200$

$n = 30$
$M = 6000$

- A wynikiem prawdziwym

PRAWDZIWY WYNIK
$N = 702$
$M = 5720,80$ zł

$n = 5$
błąd = 7000 – 5720,80 = 1279,20

$n = 10$
błąd = 5720,80 – 5200 = 520,80

$n = 30$
błąd = 6000 - 5720,80= 279,20

Błąd pomiaru

- Różnica między **szacowanym wynikiem**

$n = 5$
$M = 7000$

$n = 10$
$M = 5200$

$n = 30$
$M = 6000$

- **A wynikiem prawdziwym**

PRAWDZIWY WYNIK
$N = 702$
$M = 5720{,}80$ zł

$n = 5$
błąd = 7000 – 5720,80 = 1279,20

$n = 10$
błąd = 5720,80 – 5200 = 520,80

$n = 30$
błąd = 6000 - 5720,80= 279,20

Średni błąd pomiaru vel Standardowy błąd pomiaru

- Powtarzasz swoje losowanie dowolnej próbki *n*-elementowej X razy
- Dla każdego losowania wyliczasz błąd pomiaru
- Liczysz średnią ze wszystkich błędów

Błąd standardowy –wskazuje o ile średnio myliłbyś się w szacowaniu jakiegoś parametru, gdybyś dane badanie (prowadzone na wynikach próbki n-elementowej) powtarzał wielokrotnie.

Błąd standardowy (*SE*)

Wielkość próby *n*	Średnie wynagrodzenia wyliczone ze średnich uzyskanych w 30 analizach na próbach *n*	Różnica oszacowania między wyliczoną średnią (estymatorem) a parametrem (5720,80 zł)	Błąd standardowy (SE)
5	5630,40	90,40	1062,50
10	5771,70	50,90	464,15
30	5709,30	11,50	312,40

$$SE = \frac{odchylenie\ standardowe\ (SD)}{\sqrt{N}}$$

Im większa próba, tym większa precyzja wnioskowania

Prawo wielkich liczb Bernoulliego

- *Jeżeli z populacji X wylosuje się wiele prób (np. 100)*
- *o tej samej liczebności n*-elementowej *(np. 30)*
- *i dla każdej z tych prób (tj. prób po n = 30) obliczy się średnią arytmetyczną,*
- *to prawdopodobieństwo, że średnia arytmetyczna z tych średnich (średnia ze 100 średnich uzyskanych w badaniach po n = 30) będzie taka sama jak średnia w populacji X*
- *zbliża się do 1 wraz ze wzrostem liczebności tych próbek (czyli ze wzrostem samego n).*

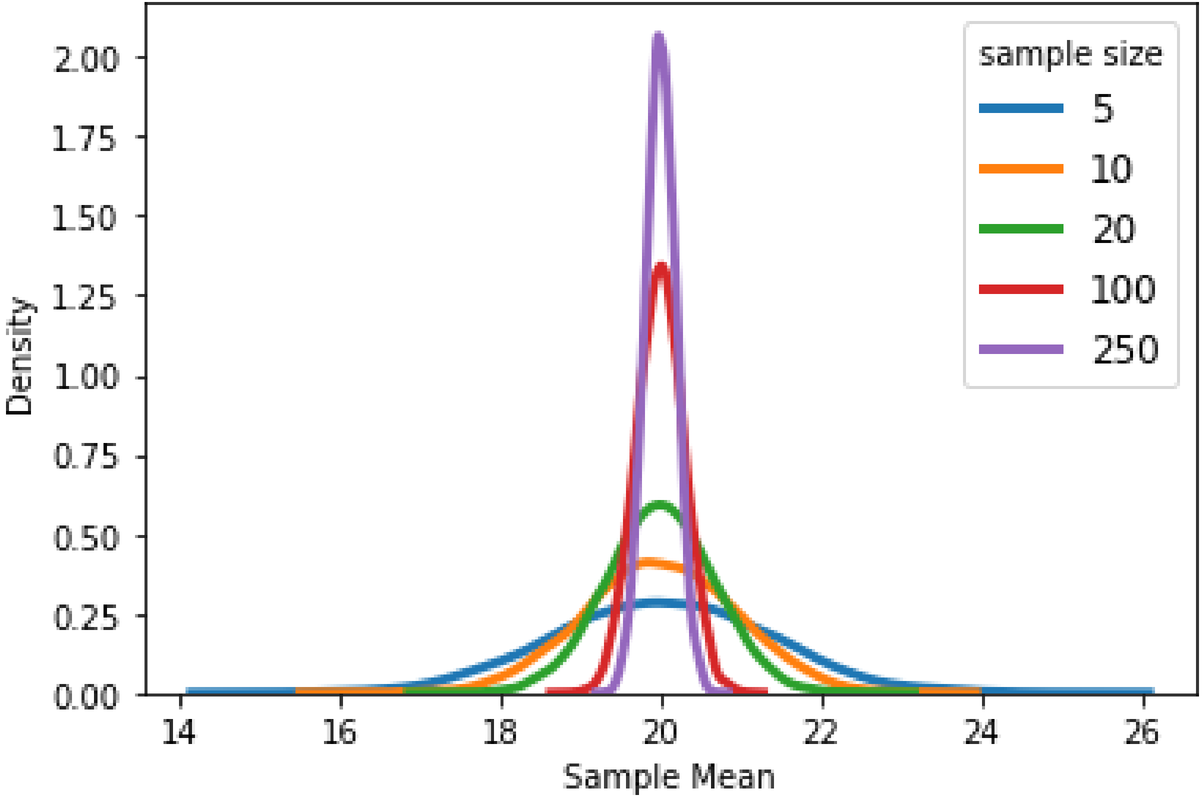
Wielkość próby *n*	Średnie wynagrodzenia wyliczone ze średnich uzyskanych w 30 analizach na próbach *n*	Różnica oszacowania między wyliczoną średnią (estymatorem) a parametrem (5720,80 zł)	Błąd standardowy (SE)
5	5630,40	90,40	1062,50
10	5771,70	50,90	464,15
30	5709,30	11,50	312,40

To właśnie dlatego w $n = 30$ uzyskano najmniejszą różnicę względem parametru i najmniejszy *SE*

- gdyby zamiast losować 30 x 30 ofert,
- wylosować 30 x 50 ofert albo x 100 ofert, o
- ostatecznie (i zgodnie z w/w prawem) błąd pomiaru zbliżałby się do zera, a uzyskana średnia – do tej prawdziwej wyliczonej w populacji.

SE jest tym większy

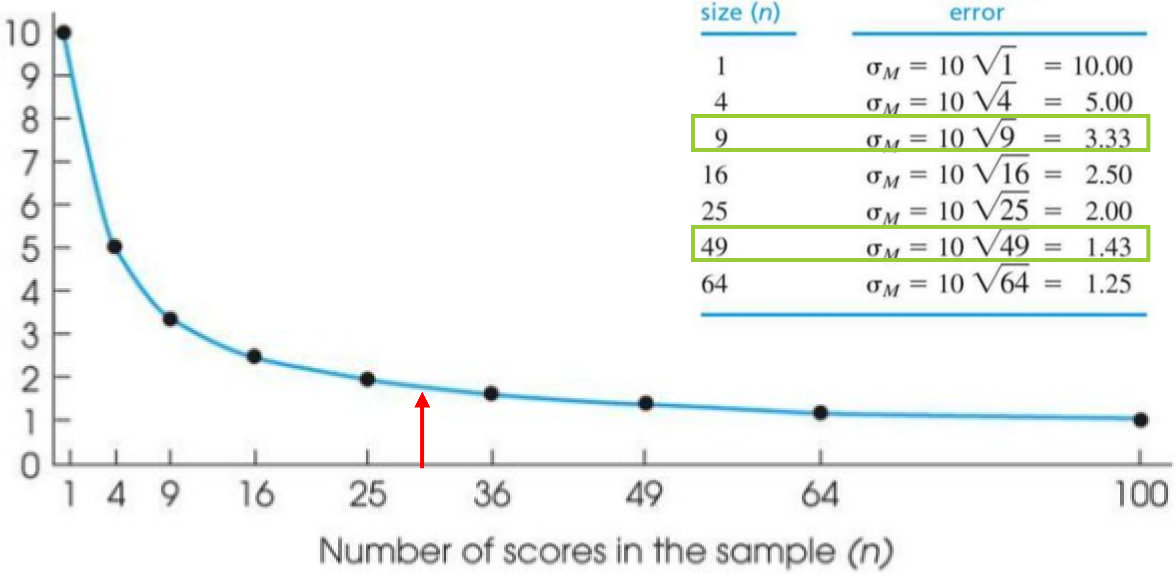
- Im ↓ próba
- Im ↑ rozrzut wyników wokół średniej



Ile to jest wystarczająco duża próba?

Określenie dokładnej relacji między **precyzją szacowania** a **wielkością próby**, wymaga wiedzy o SE dla prób o różnej wielkości

Sample size (n)	Standard error
1	$\sigma_M = 10\sqrt{1} = 10.00$
4	$\sigma_M = 10\sqrt{4} = 5.00$
9	$\sigma_M = 10\sqrt{9} = 3.33$
16	$\sigma_M = 10\sqrt{16} = 2.50$
25	$\sigma_M = 10\sqrt{25} = 2.00$
49	$\sigma_M = 10\sqrt{49} = 1.43$
64	$\sigma_M = 10\sqrt{64} = 1.25$



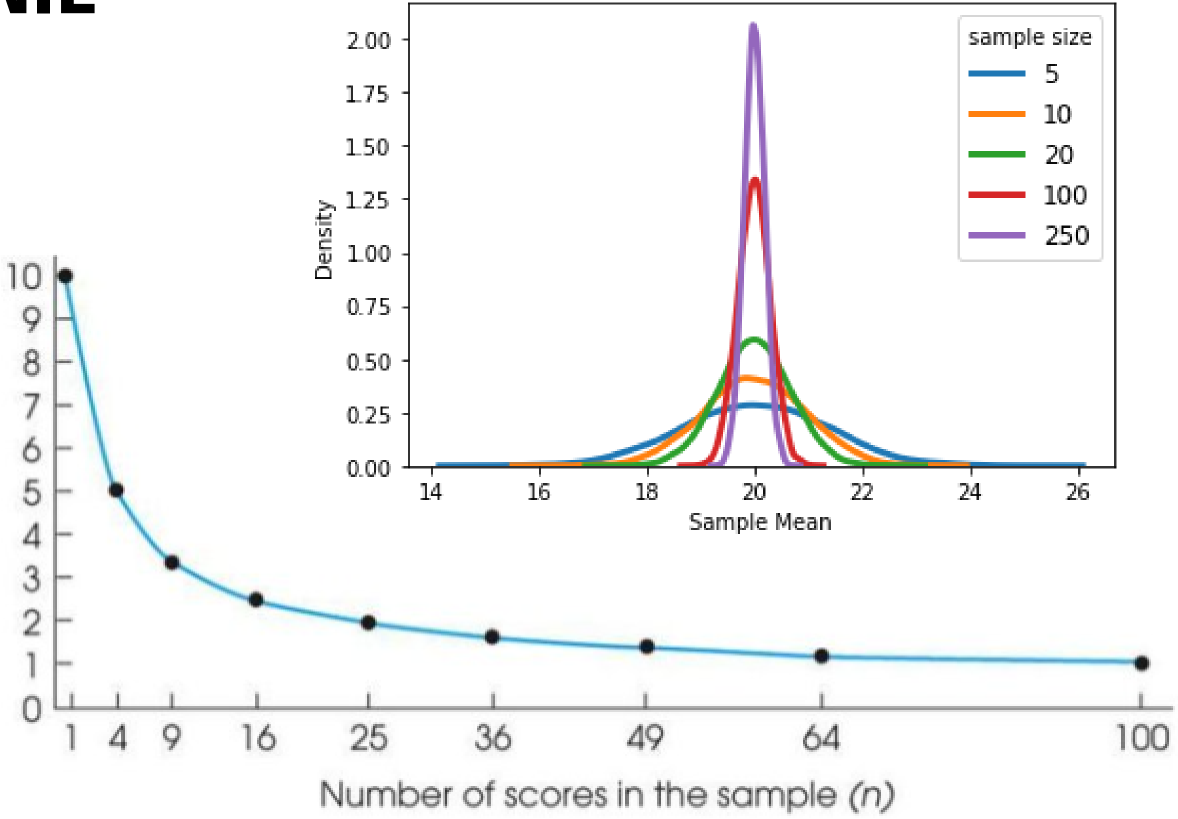
to zależy....

- Jakie zjawisko badasz
- Co chcesz w badaniu uzyskać
- W jaki sposób będą analizowane dane
- Z jakich populacji te dane pochodzą
- Jaki mają rozkład

PODSUMOWANIE

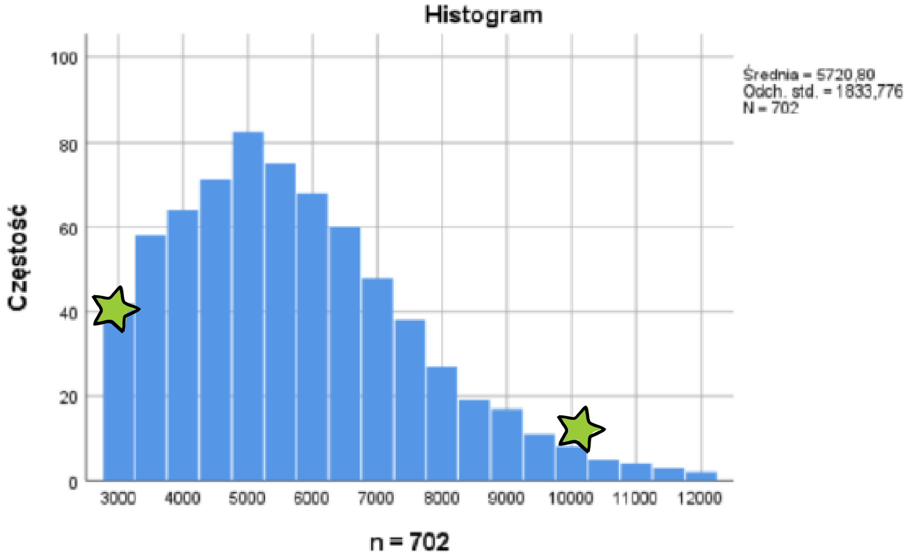
SE jest tym większy

- Im ↓ próba
- Im ↑ rozrzut wyników wokół średniej



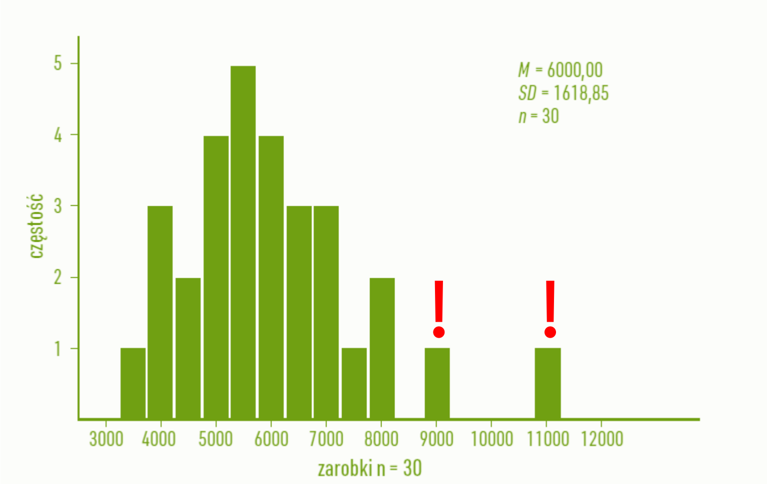
Rozkład danych - histogram

- Histogram – pozwala na graficzne przedstawienie częstości występowania danej wartości w zbiorze.



Obserwacje odstające (*outliers*)

Obserwacje odstające to takie, które w znaczący sposób odbiegają od pozostałych

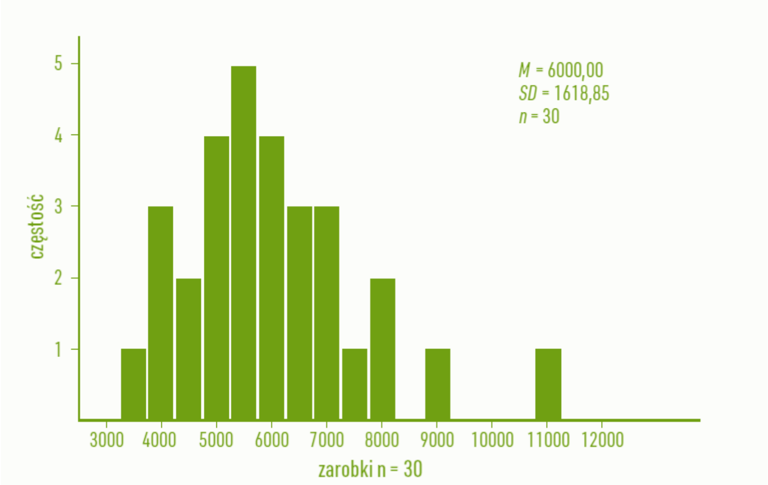


Takie skrajne wartości powodują, iż szacowane wyniki mogę nie odzwierciedlać tych realnych, utrudniając w ten sposób dalszą analizę

Rozkład

- W populacji
- W próbie
- Z próby (próbkowania)

Rozkład empiryczny (rozkład **w** próbie)



- Rozkład danych surowych
- czyli rozkład wyników pojedynczego badania, **który jednocześnie jest jedynym**, którego właściwości można określić
- Właściwości rozkładu:
  - Kształt
  - Średnia
  - Odchylenie
  - Skośność itp.

Rozkład w populacji

- rozkład wyników w całej populacji (możliwy do poznania w pełni, o ile jest ona skończona i w całości przebadana);
- zwykle jest to jednak rozkład o nieznanym kształcie i nieznanych parametrach, którego odbiciem stają się wyniki przeprowadzonego badania (próby)
- np. rozkład proponowanych pensji z 702 ofert

Rozkład z próby (rozkład próbkowania)

- jest **rozkładem teoretycznym** jakiejś statystyki / estymatora (np. średniej) uzyskanej z większej liczby prób pobranych z tej samej populacji.
- Oddaje częstotliwość występowania rożnych wartości statystyki wyliczonych na podstawie hipotetycznych kolejnych prob.
- Pozwala określić **precyzję pomiaru**, czyli tak naprawdę zależność między wartością rzeczywistą w populacji a zmierzoną w próbie
- np. byłby to rozkład średnich proponowanych pensji z wielokrotnie powtarzanych badań na próbach n = 30.

Ważne

Rozkład z próby (próbkowania) stanowi podstawę we wnioskowaniu statystycznym (podejście klasyczne).

Analizując poszczególne testy statystyczne, założenie o rozkładzie normalnym dotyczy rozkładu próbkowania!

Rozkład próbkowania a precyzja pomiaru

Najlepszą miarą precyzji pomiaru jest SE (średni błąd pomiaru)

W praktyce jego wartości nie da się określić:

- Nieznana jest wartość prawdziwego parametru (częściej nasza populacja jest hipotetyczna np. studenci, Polacy itp.
- Zazwyczaj wykonuje się 1 badanie a nie całą serię!

Z odsieczą przychodzi **szacunkowy błąd standardowy**

Szacunkowy błąd standardowy

- Opiera się na wynikach w próbie
  - Wariancji, z której pierwiastek da Ci odchylenie standardowe (SD)
  - Wariancja to kwadrat odległości każdej osoby od średniej
  - SD – to pierwiastek z wariancji
- Opierając się na wariancji i liczebności próby można obliczyć o ile w każdym pomiarze można by było się pomylić

Obliczanie

Gdybyś nie znał parametrów charakteryzujących populację 702 ofert dotyczących pracy na stanowisku statystyka, mógłbyś wykorzystać dane z próby n = 30 do ich oszacowania.

Wyliczona na ich podstawie średnia kwota wyniosła

M = 6000 zł, wariancja = 2 620 675,32 SD = 1618,85

$$Szacunkowy\ SE = \sqrt{\frac{wariancja}{liczba\ obserwacji}} = 295{,}56\ zł$$

O taką kwotę średnio pomyliłbyś się szacując wynik na 30 el. próbie

To ile mogę zarabiać?

Aby określić „widełki zarobków”, od średniej odejmujesz i dodajesz SD (pierwiastek z wariancji)

SD = 1618,85

M = 4480,03

W przypadku idealnego rozkładu normalnego **68,3%** zarobków mieści się między **4311,15 zł do 7618,85 zł**

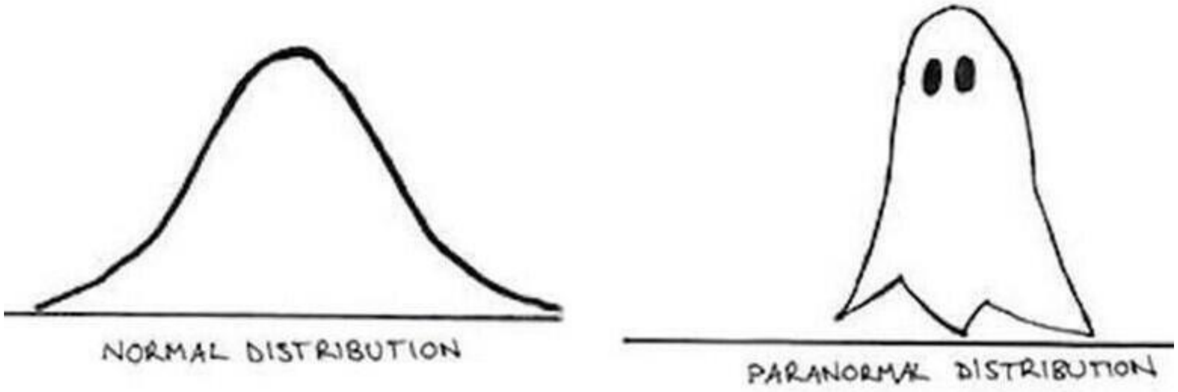
Odchylenie standardowe

- jest miarą opisową dotyczącą rozkładu danych nt. jakiejś zmiennej (np. kwot wynagrodzenia) w pojedynczej próbie.
- W przykładzie powyżej jest to właśnie liczba 1618,85, która obrazuje kwotę zmienności po obu stronach średniej (czyli in plus i in minus),
- dlatego często zapisuje się je w parze, tj. średnia ± odchylenie (w przykładzie powyżej: 6000 ± 1618,75 zł).

Rozkład

SD i wariancja → rozkładu w próbie

Błąd standardowy (SE) → rozkładu próbkowania (rozkładu z próby)



Rozkład próbkowania

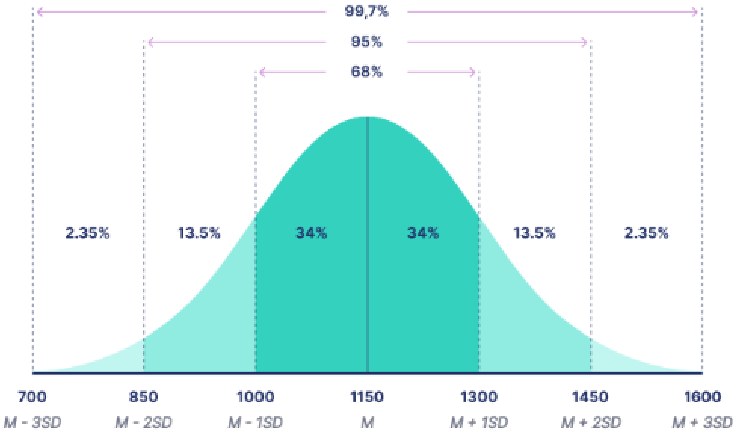
- Opiera się o Centralne Twierdzenie Graniczne
- Przyjmuje kształt rozkładu normalnego (krzywej Gaussa)

WAŻNE!

- Klasyczny rozkład normalny ma niezmienne właściwości matematyczne!
- Dzięki czemu istnieje możliwość ustalenia prawdopodobieństwa występowania dowolnego obszaru pod krzywą

Rozkład normalny

- Ma centralnie położoną średnią
- Jest symetryczny
- Istnieje 50% prawdopodobieństwo na uzyskanie wartości mniejszej/większej od średniej



Standaryzacja Z

- Dane z ofert pracy były liczone w zł
- Aby wyliczyć prawdopodobieństwo występowania danej kwoty np. zarobków w kwocie 8000 zł lub 10000 zł
- Należy rozkład z naszej próby (n = 30) wystandaryzować
- Czyli zamienić ze złotówek na *zety czyli wartości standaryzowane*

Standaryzacja Z - przykład

Nasz rozkład ma

M = 6000 zł

SD = 1618,85

Rozkład normalny

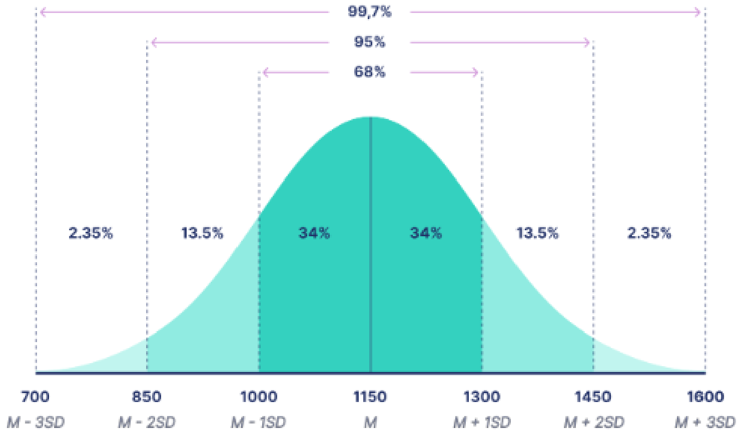
M = 0

SD = 1

Aby wyliczyć prawdopodobieństwo zarobków na poziomie 8-10k

Trzeba znaleźć adekwatną wartość Z

Standaryzacja Z - przykład



Nasz rozkład ma
M = 6000 zł
SD = 1618,85

Rozkład normalny
M = 0
SD = 1

M +/- SD = 4311,15–7618,85 – pensja z tego przedziału mieści się w +/-1 SD – oznacza, że masz 68% szans na otrzymanie takiego wynagrodzenia

7618,85 + 1618,85 (1SD) = 9237,70
pensja w zakresie 7618,85 - 9237,70 masz 13,5% szans
pensja 10k mieści się w 3SD – masz już tylko 2,35% szans

Standaryzacja Z - przykład

- Należy użyć wzoru na proporcję między różnicą wskazanej kwoty i średniej z próby przez odchylenie standardowe
- Dla kwoty 8k i 10k

$$z = \frac{8000 - 6000\,(M)}{1618{,}85} = \frac{2000}{1618{,}58} = 1{,}24$$

$$z = \frac{10000 - 6000\,(M)}{1618{,}85} = \frac{4000}{1618{,}58} = 2{,}48$$

Standaryzacja Z - przykład

Szukane prawdopodobieństwo dla
z w zakresie 1,24 – 2,48

0,107- 0,007
Czyli między 10,7% a 0,07%

z	p_z	z	p_z	z	p_z	z	p_z	z	p_z	z	p_z	z	p_z
0	0,5000	0,5	0,3085	1	0,1587	1,5	0,0668	2	0,0228	2,5	0,0062	3	0,0013
0,01	0,4960	0,51	0,3050	1,01	0,1562	1,51	0,0655	2,01	0,0222	2,51	0,0060	3,01	0,0013
0,02	0,4920	0,52	0,3015	1,02	0,1539	1,52	0,0643	2,02	0,0217	2,52	0,0059	3,02	0,0013
0,03	0,4880	0,53	0,2981	1,03	0,1515	1,53	0,0630	2,03	0,0212	2,53	0,0057	3,03	0,0012
0,04	0,4840	0,54	0,2946	1,04	0,1492	1,54	0,0618	2,04	0,0207	2,54	0,0055	3,04	0,0012
0,05	0,4801	0,55	0,2912	1,05	0,1469	1,55	0,0606	2,05	0,0202	2,55	0,0054	3,05	0,0011
0,06	0,4761	0,56	0,2877	1,06	0,1446	1,56	0,0594	2,06	0,0197	2,56	0,0052	3,06	0,0011
0,07	0,4721	0,57	0,2843	1,07	0,1423	1,57	0,0582	2,07	0,0192	2,57	0,0051	3,07	0,0011
0,08	0,4681	0,58	0,2810	1,08	0,1401	1,58	0,0571	2,08	0,0188	2,58	0,0049	3,08	0,0010
0,09	0,4641	0,59	0,2776	1,09	0,1379	1,59	0,0559	2,09	0,0183	2,59	0,0048	3,09	0,0010
0,1	0,4602	0,6	0,2743	1,1	0,1357	1,6	0,0548	2,1	0,0179	2,6	0,0047	3,1	0,0010
0,11	0,4562	0,61	0,2709	1,11	0,1335	1,61	0,0537	2,11	0,0174	2,61	0,0045	3,11	0,0009
0,12	0,4522	0,62	0,2676	1,12	0,1314	1,62	0,0526	2,12	0,0170	2,62	0,0044	3,12	0,0009
0,13	0,4483	0,63	0,2643	1,13	0,1292	1,63	0,0516	2,13	0,0166	2,63	0,0043	3,13	0,0009
0,14	0,4443	0,64	0,2611	1,14	0,1271	1,64	0,0505	2,14	0,0162	2,64	0,0041	3,14	0,0008
0,15	0,4404	0,65	0,2578	1,15	0,1251	1,65	0,0495	2,15	0,0158	2,65	0,0040	3,15	0,0008
0,16	0,4364	0,66	0,2546	1,16	0,1230	1,66	0,0485	2,16	0,0154	2,66	0,0039	3,16	0,0008
0,17	0,4325	0,67	0,2514	1,17	0,1210	1,67	0,0475	2,17	0,0150	2,67	0,0038	3,17	0,0008
0,18	0,4286	0,68	0,2483	1,18	0,1190	1,68	0,0465	2,18	0,0146	2,68	0,0037	3,18	0,0007
0,19	0,4247	0,69	0,2451	1,19	0,1170	1,69	0,0455	2,19	0,0143	2,69	0,0036	3,19	0,0007
0,2	0,4207	0,7	0,2420	1,2	0,1151	1,7	0,0446	2,2	0,0139	2,7	0,0035	3,2	0,0007
0,21	0,4168	0,71	0,2389	1,21	0,1131	1,71	0,0436	2,21	0,0136	2,71	0,0034	3,21	0,0007
0,22	0,4129	0,72	0,2358	1,22	0,1112	1,72	0,0427	2,22	0,0132	2,72	0,0033	3,22	0,0006
0,23	0,4090	0,73	0,2327	1,23	0,1093	1,73	0,0418	2,23	0,0129	2,73	0,0032	3,23	0,0006
0,24	0,4052	0,74	0,2296	1,24	0,1075	1,74	0,0409	2,24	0,0125	2,74	0,0031	3,24	0,0006
0,25	0,4013	0,75	0,2266	1,25	0,1056	1,75	0,0401	2,25	0,0122	2,75	0,0030	3,3	0,0005
0,26	0,3974	0,76	0,2236	1,26	0,1038	1,76	0,0392	2,26	0,0119	2,76	0,0029	3,4	0,0003
0,27	0,3936	0,77	0,2206	1,27	0,1020	1,77	0,0384	2,27	0,0116	2,77	0,0028	3,5	0,0002
0,28	0,3897	0,78	0,2177	1,28	0,1003	1,78	0,0375	2,28	0,0113	2,78	0,0027	3,6	0,0002
0,29	0,3859	0,79	0,2148	1,29	0,0985	1,79	0,0367	2,29	0,0110	2,79	0,0026	3,7	0,0001
0,3	0,3821	0,8	0,2119	1,3	0,0968	1,8	0,0359	2,3	0,0107	2,8	0,0026		
0,31	0,3783	0,81	0,2090	1,31	0,0951	1,81	0,0351	2,31	0,0104	2,81	0,0025		
0,32	0,3745	0,82	0,2061	1,32	0,0934	1,82	0,0344	2,32	0,0102	2,82	0,0023		
0,33	0,3707	0,83	0,2033	1,33	0,0918	1,83	0,0336	2,33	0,0099	2,83	0,0023		
0,34	0,3669	0,84	0,2005	1,34	0,0901	1,84	0,0329	2,34	0,0096	2,84	0,0023		
0,35	0,3632	0,85	0,1977	1,35	0,0885	1,85	0,0322	2,35	0,0094	2,85	0,0022		
0,36	0,3594	0,86	0,1949	1,36	0,0869	1,86	0,0314	2,36	0,0091	2,86	0,0021		
0,37	0,3557	0,87	0,1922	1,37	0,0853	1,87	0,0307	2,37	0,0089	2,87	0,0021		
0,38	0,3520	0,88	0,1894	1,38	0,0838	1,88	0,0301	2,38	0,0087	2,88	0,0020		
0,39	0,3483	0,89	0,1867	1,39	0,0823	1,89	0,0294	2,39	0,0084	2,89	0,0019		
0,4	0,3446	0,9	0,1841	1,4	0,0808	1,9	0,0287	2,4	0,0082	2,9	0,0019		
0,41	0,3409	0,91	0,1814	1,41	0,0793	1,91	0,0281	2,41	0,0080	2,91	0,0018		
0,42	0,3372	0,92	0,1788	1,42	0,0778	1,92	0,0274	2,42	0,0078	2,92	0,0018		
0,43	0,3336	0,93	0,1762	1,43	0,0764	1,93	0,0268	2,43	0,0075	2,93	0,0017		
0,44	0,3300	0,94	0,1736	1,44	0,0749	1,94	0,0262	2,44	0,0073	2,94	0,0016		
0,45	0,3264	0,95	0,1711	1,45	0,0735	1,95	0,0256	2,45	0,0071	2,95	0,0016		
0,46	0,3228	0,96	0,1685	1,46	0,0721	1,96	0,0250	2,46	0,0069	2,96	0,0015		
0,47	0,3192	0,97	0,1660	1,47	0,0708	1,97	0,0244	2,47	0,0068	2,97	0,0015		
0,48	0,3156	0,98	0,1635	1,48	0,0694	1,98	0,0239	2,48	0,0066	2,98	0,0014		
0,49	0,3121	0,99	0,1611	1,49	0,0681	1,99	0,0233	2,49	0,0064	2,99	0,0014		

Testy statystyczne

Powyższe przykłady zabiegów (tj. np. standaryzacja rozkładu normalnego) i powstałe dzięki nim uniwersalne charakterystyki rozkładów pozwoliły na opracowanie **testów statystycznych**, czyli wzorów umożliwiających określenie prawdopodobieństwa występowania poszczególnych wyników, np. różnicy pomiędzy średnimi w jakiś grupach czy miary związku między dwiema zmiennymi).

To prawdopodobieństwo pozwala natomiast określić, czy dany wynik ma znaczenie statystyczne, czy nie!

Typy testów statystycznych

parametryczne

- Dokładniejsze
- Określone założenia

nieparametryczne

- Mniej rygorystyczne założenia

Centralne Twierdzenie Graniczne

- Obok Prawa Wielkich Liczb
- CTG – stanowi podstawę stosowania min testów parametrycznych
- A brzmi:

Niezależnie od kształtu rozkładu w populacji, przy losowych i niezależnych pomiarach dokonanych w jej obrębie,
rozkład średnich z tych prób będzie zbliżał się do normalnego

– tym bardziej, im więcej zbierzemy obserwacji.

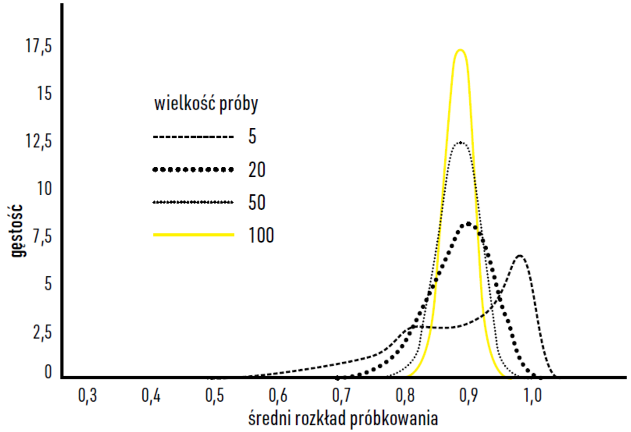
Centralne Twierdzenie Graniczne

Opierając na przykładzie z liczeniem średniej pensji ustalanej na podstawie ofert pracy, powyższe prawo można zastosować następująco:

- *Niezależnie od kształtu rozkładu w populacji*, czyli niezależnie od rozkładu wszystkich wynagrodzeń z 702 ofert (populacja),
- *przy losowych i niezależnych pomiarach dokonanych w jej obrębie*, czyli właśnie przy losowaniu np. 30 ofert (próba) z danej populacji N = 702 (o obojętnie jakim rozkładzie) i badaniu średniej pensji
- *rozkład średnich z tych prób będzie zbliżał się do normalnego* – czyli rozkład średnich (rozkład próbkowania), a nie rozkład danych!
- I UWAGA! NAJWAŻNIEJSZA CZĘŚĆ: *tym bardziej, im więcej zbierzemy obserwacji!* – czyli im więcej obserwacji liczyłaby próba, tym większe prawdopodobieństwo, że uzyskane na jej podstawie średnie (wynikające z kolejnych pomiarów w próbach o tej samej wielkości) przyjęłyby rozkład normalny.

Podsumowując

należy sobie uświadomić, że wraz ze wzrostem liczebności próby do rozkładu normalnego upodabnia się rozkład nie badanej zmiennej, a pewna jej funkcja, która wykorzystana jest do budowy testu parametrycznego



Estymacja przedziałowa

- Wcześniej wspominaliśmy o estymacji punktowej, wtedy szacowaliśmy szansę na otrzymanie konkretnej wielkości wynagrodzenia
- Estymacja przedziałowa pozwala na szacowanie „widełek”, jesteśmy w stanie znaleźć przedział do którego wartość poszukiwana może należeć
- Ten przedział nazywa się **przedziałem ufności**

Poziom ufności

- Wyznaczanie prawdopodobieństwa dla określonych wartości standardowego rozkładu normalnego pozwala na dokonanie estymacji przedziałowej.
- Procedura ta pomaga statystykom odpowiedzieć na pytanie, na ile uzyskany wynik jest godny zaufania
- Standaryzacja normalnego rozkładu prawdopodobieństwa umożliwia dokładne oszacowanie poziomu takiego „zaufania”, który w statystyce nazywany jest nie inaczej, jak… **poziomem ufności!**
- **Najczęściej przyjmuje się jego wartość na 95%, ale zdarza się 90 lub 99%**

Przedział ufności (CI, *Confidnece Interval*)

- Ustalany na podstawie poziomu ufności np. 95%
- Przy poziomie ufności 95% wartość mieści się między +/-1,96 SD rozkładu normalnego
- Żeby wyliczyć wartość CI potrzebujesz informację o średniej (*M) i błędzie standardowym*.
- Dla prób n > 30

średnia (*M*) +/– (1,96 * błąd standardowy średniej)

Przykład

dla przykładu z wyliczaniem średnich oferowanych zarobków na stanowisku statystyka na podstawie n = 30 ofert:

6000 (*M*) – 579,30 (błąd standardowy średniej) = 5420,70 zł

6000 + 579,30 = 6579,30 zł

$$SE = \frac{odchylenie\ standardowe\ (SD)}{\sqrt{N}}$$

Interpretacja

- Opiera się na założeniu, że wyliczony, pojedynczy przedział ufności jest jednym z nieskończonej ilości możliwych do wyliczenia przedziałów.
- W związku z tym X% (np. 95%) przedział ufności określa prawdopodobieństwo tego, że wyliczony zakres wartości jest jednym z tych X% (np. 95%), w których rzeczywiście znajduje się szacowany parametr.
- Przykładowo, jeśli dane badanie zostanie powtórzone 100 razy, to 95%, czyli 95 wyliczonych z nich przedziałów ufności (z których każdy będzie jednak nieco inny, bo choć liczony jest dla tej samej liczebności próby, to opiera się na innych danych!) będzie zawierać rzeczywistą wartość parametru (czyli np. prawdziwą średnią), a 5% z nich, czyli 5 – nie.

Na szczęście!

Przedziały ufności wyliczy za Ciebie program statystyczny!

Ale TY powinieneś wiedzieć co to oznacza!

Testowanie hipotez

Funkcjonują dwa rodzaje hipotez:

1. Hipoteza zerowa – mówiąca o braku efektu
2. Hipoteza alternatywna – mówiąca o występowaniu efektu

Np.

H_0 zarobki są równe 5000 zł (nie występuje żadna różnica)

H_1 zarobki są wyższe niż 5000 zł (nie występuje żadna różnica)

Hipoteza statystyczna a badawcza

Różnica między hipotezą badawczą a statystyczną polega na tym, że najczęściej zakłada się bardzo konkretne efekty (hipotezy kierunkowe), podczas gdy w hipotezach statystycznych stawianych dla testów dwustronnych mówi się tylko o ogólnych zależnościach

(jest różnica vs nie ma różnicy, jest związek vs nie ma związku, jest wpływ vs nie ma wpływu itp.) (hipotezy niekierunkowe).

Jeszcze raz

Założenie

Influencerzy zarabiają więcej niż statystycy

Hipoteza badawcza
Średnia kwota zarobków influencerów jest wyższa niż średnia kwota zarobków statystyków

H0 Średnia kwota wynagrodzenia oferowana internetowym influencerom jest taka sama jak średnia kwota oferowana statystykom (rożnica = 0)

H1 Średnia kwota wynagrodzenia oferowana internetowym influencerom jest rożna od średniej kwoty oferowanej statystykom (rożnica ≠ 0)

Testowanie hipotez

Uwaga NIEINTUICYJNE!

Weryfikacja zachodzi

nie poprzez określenie prawdopodobieństwa prawdziwości hipotezy alternatywnej, czyli tej stwierdzającej, że dany efekt występuje,

tylko poprzez odrzucenie hipotezy zerowej, mówiącej o braku efektu.

Hipoteza zerowa

- Jest „nadrzędna" nad alternatywną, bo właśnie ona jest testowana
- W wyniku analizy można TYLKO wyciągnąć wniosek *„że nie ma podstaw do odrzucenia hipotezy zerowej"*

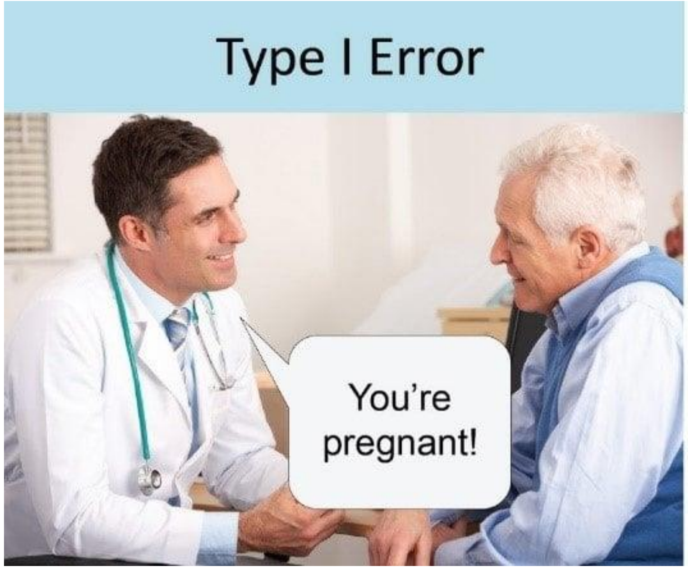
- Przyjmowanie hipotezy – jest określeniem potocznym bo odrzucenie H0 nie jest dowodem na nieistnienie określonego efektu, może być wynikową specyfiki grupy, błędu pomiaru itp.

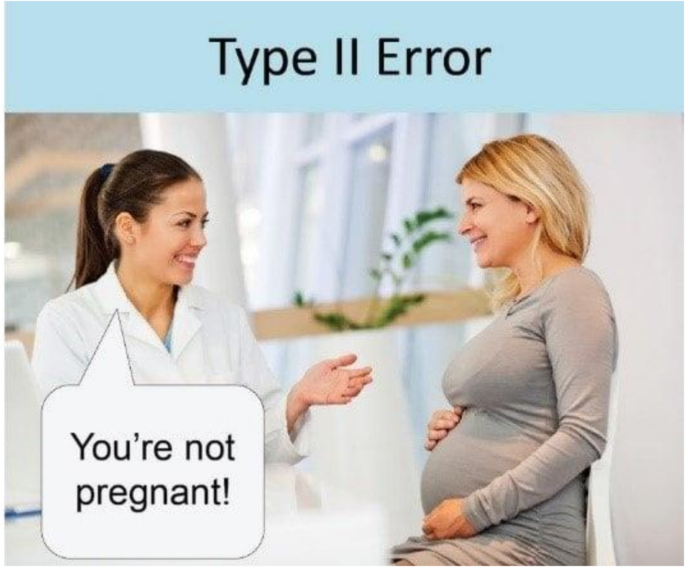
Błąd I i II rodzaju

Błąd I rodzaju (określany mianem α – alfa) polega na odrzuceniu hipotezy zerowej wtedy, kiedy realnie jest prawdziwa (czyli na stwierdzeniu efektu, w sytuacji, kiedy on nie występuje, np. stwierdzeniu różnic w zarobkach statystyków i influencerów, podczas gdy takich nie ma).

Błąd II rodzaju (określany mianem β – beta) polega na przyjęciu hipotezy zerowej, która w rzeczywistości jest fałszywa (czyli na stwierdzeniu braku efektu, który w rzeczywistości istnieje, np. stwierdzeniu, że zarobki statystyków i influencerów nie różnią się, podczas gdy tak naprawdę różnice istnieją).

Błąd I i II rodzaju





Poziom istotności (α) i poziom ufności

- **poziom istotności (α)** rozumiany jako prawdopodobieństwo / ryzyko popełnienia **błędu I rodzaju** (tego, który polegał na odrzuceniu prawdziwej hipotezy zerowej). Jego wartość jest zależna od przyjętego poziomu ufności, ale zwykle uznaje się, że wynosi ona 5% (lub 1%, a czasem nawet jeszcze mniej).
- **Poziom ufności** (który najczęściej umownie określa się na 95%, ale można również na 99%, czy 99,9%) i w ogóle ufność, rozumiane zgodnie z pojęciem częstościowym, jako procent **przedziałów ufności** (np. dla poziomu 95% jest ich 95%), które szacowane na podstawie nieskończonej liczby powtórzeń danego badania, zawierają prawdziwą wartość estymowanego parametru (np. średnia wartość proponowanych zarobków w ogłoszeniach o pracę

Przedziały ufności

- mają na celu zobrazowanie, jak godny zaufania jest uzyskany w badaniu estymator (np. obliczona średnia) i wiążą się z błędem standardowym.
- WAŻNE
  uzyskany w badaniu przedział nie wskazuje ani na prawdopodobieństwo, ani procent twojej pewności, mówiące o tym, że prawdziwy parametr (np. rzeczywista średnia, którą szacuje się za pomocą estymatora) leży w jego granicach.
- ów zakres oznacza, że **jest on jednym z 95% (w przypadku takiego właśnie poziom ufności) tych, które zawierają poszukiwaną wartość (czyli parametr)**.

Przykład średnia wynagrodzeń

n = 30 M = 6000 95% CI = 5420,70-6679,30

- wskazuje to na fakt, że przy wielokrotnym powtarzaniu takiego badaniu i uzyskaniu innych zakresow ten przedstawiony powyżej, podobnie jak reszta 95% tak wyliczonych przedziałow, zawiera realną średnią kwotę proponowanych pensji.
- W tym powyższym wypadku możesz uznać zatem, że średnia leży właśnie gdzieś pomiędzy 5420,70 a 6679,30, ale pamiętaj, że wykonując chociażby drugi raz identyczne badanie

  n = 30 M = 5183,33 95% CI = 4582,55-5784,11

prawdziwa średnia leżałaby także i w tym drugim zakresie.

Poziom istotności (α)

- im większa jego wartość (czyli np. nie 5%, a 10%), tym większe prawdopodobieństwo popełnienia błędu I rodzaju i mniejszy poziom ufności (wynosi on wówczas już nie 95%, a 90%)

Moc testu

- Wrażliwość metody na wykrywanie określonego efektu związana jest z pojęciem **mocy testu statystycznego**.
- Prawdopodobieństwo uniknięcia błędu II rodzaju (tego, który polegał na nieodrzuceniu fałszywej hipotezy zerowej), a zatem im większa moc testu, tym większa jego zdolność do odrzucenia hipotezy zerowej (o ile jest fałszywa!).
- Moc testu statystycznego zależy od poziomu istotności α, ale nie tylko od niego!

Moc testu

- Wielkości/ siły testowanego efektu w populacji – jeśli efekt jest znaczący to trudniej go nie zauważyć
- Liczebność próby – zmniejsza wielkość błędu standardowego i tym samym zawęża przedziały ufności

- DODATKOWO
- Moc testu jest różna dla różnych testów statystycznych np. parametryczne mają większą czułość

Moc testu – w praktyce

- Polega na zapewnieniu sobie wystarczająco dużej próby badawczej
- UWAGA jeżeli zakładany efekt jest mały, będzie potrzebna większa próba niż n = 30

- Wiedząc jakiej siły efektu się spodziewasz (np. z przeglądu literatury) oraz jaką analizę, na ilu grupach chcesz przeprowadzić – możesz wykorzystać kalkulator np. g*power który wskaże jak duża powinna być grupa

Czy grupa może być za duża?

- Wyobraź sobie, że badając 1000 (n = 1000) statystyków i 1000 (n = 1000) influencerów otrzymałeś średnie wysokości pensji równe kolejno 5800 zł i 5882 zł (ta druga pensja wydaje się niska, ale przypuśćmy, że zbadałeś wielu „świeżaków” w tej branży, a co za tym idzie, ich zarobki nie odbiegają specjalnie od średniej krajowej).
- Wykonując test statystyczny, otrzymujesz wynik, który sugeruje odrzucenie hipotezy zerowej i sprowadza się do wniosku, że statystycy i influencerzy różnią się istotnie pod względem średnich zarobków. Ale czy rzeczywiście ta różnica (dokładnie 82 zł)ma znaczenie? Czy miałaby znaczenie dla ciebie, gdybyś szukał pracy?
- Takie wyniki są właśnie konsekwencją nadmiernej czułości testu, a co za tym idzie, zwiększonej zdolności do wykrywania efektów o nawet minimalnej sile.

Statystyka testowa

- każdy test, a właściwie jego formuła, która zostaje wykorzystana do obliczania wyników, **opiera się na pewnym rozkładzie prawdopodobieństwa ustalonym na podstawie wzorów matematycznych i probabilistyki**.
- Ten ROZKŁAD – jest rozkładem statystyki testowej
- Statystyka testowa to taka **zmienna losowa** (czyli wartość, która może się zmieniać w sposób dowolny, w zależności od badania), która każdorazowo obliczana na podstawie wyników uzyskanych w próbie, pozwala podjąć decyzję o odrzuceniu lub nie hipotezy zerowej.
- Jej idea wynika z tego, że nie da się na podstawie surowych danych ocenić ich istotności, a **dzięki takiej właśnie uniwersalnej transformacji (polegającej na sprowadzeniu wyniku analizy wykonywanej na dowolnej liczbie dowolnych danych w dowolnych jednostkach do jednej wartości), proces weryfikacji jest możliwy**.

Statystyka testowa

- wykonując test statystyczny, określa się szansę (prawdopodobieństwo) wystąpienia wartości danej statystyki testowej, która wyliczana jest na podstawie wyników uzyskanych w próbie.
- Co ważne, rozkład ten odnosi się do hipotezy zerowej, czyli przedstawia prawdopodobieństwo wystąpienia określonych wartości statystyki przy założeniu, że H0 jest prawdziwa (czyli kiedy brak jest efektu)
- Rozkład statystyki testowej to zatem nic innego, jak rozkład jej wartości (obliczonych na podstawie wyników w próbie) i przypisanego im prawdopodobieństwa wystąpienia, ale przy założeniu – co ważne – że stawiana w danej analizie hipoteza zerowa jest prawdziwa.

Wartość *p* i wartość krytyczna

- poziom istotności to $\alpha = 0{,}05$
- To kryterium oznacza, że jeśli w toku analizy danych w jakiejś próbie prawdopodobieństwo wylosowania określonej statystyki z populacji, w której prawdziwa jest hipoteza zerowa, będzie mniejsze niż 5% (< 0,05), hipotezę tę można będzie odrzucić.
- Prawdopodobieństwo, o którym mowa, a które porównujesz do ustalonego poziomu α i na podstawie którego wyciągasz takie wnioski (tj. odrzucasz lub nie hipotezę zerową), to właśnie słynna wartość *p* (p od *probability*)!

Wartość *p* i wartość krytyczna

- Poziom istotności $\alpha = 0{,}05$ oznacza dopuszczalne 5% ryzyka na odrzucenie prawdziwej hipotezy zerowej (popełnienie błędu I rodzaju), co przy okazji determinuje 95% poziom ufności; jest to wartość, która stanowi konwencjonalne kryterium, do którego porównuje się wyliczone *p*, a następnie podejmuje decyzję o odrzuceniu lub nie hipotezy zerowej
- Poziom prawdopodobieństwa *p* (często w programach statystycznych określany jako istotność) = 0,050 oznacza, że przy założeniu, iż hipoteza zerowa jest prawdziwa (testowana zależność nie istnieje), szansa wylosowania z populacji takiej próby, dla której statystyka testowa wynosi co najmniej tyle, ile uzyskana w badaniu wartość, równa się 5% (czyli zgodnie z kryterium istotności wystarczająco mało!).
- oznacza to ponadto, że statystyka testowa osiągnęła **wartość krytyczną**, od której przyjmuje się, że wynik jest istotny!

Wartość *p* i wartość krytyczna

Z powyższego wynika, że wartości istotności, czyli prawdopodobieństwa *p* wyliczanego dla danej statystyki testowej:

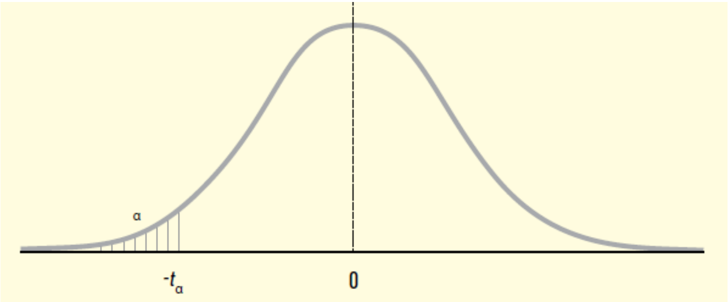
1. spełniające kryterium $<$ (mniejsze od) 0,050 (bo $\alpha = 0,05$), są istotne!
2. niespełniające powyższego kryterium, a co za tym idzie $>$ (większe od) 0,050 (bo $\alpha = 0,05$), są nieistotne!

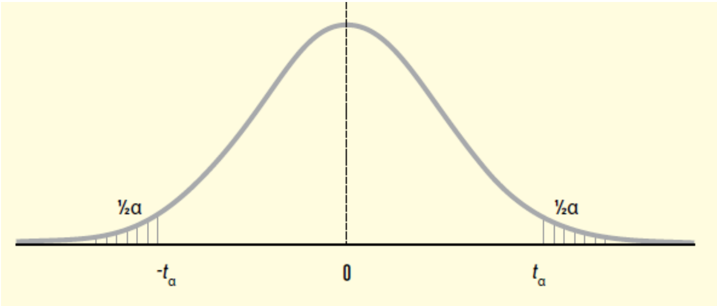
Wartość krytyczna

- Wartość krytyczna to wielkość statystyki testowej, dla której przypisane prawdopodobieństwo (wartość *p*) spełnia założone kryterium istotności (najczęściej $\alpha = 0,05$, ale może być też $\alpha = 0,01$ lub 0,001).
- Każdy rozkład statystyki testowej ma tzw. obszar krytyczny, którego jedną z granic wyznacza właśnie wartość krytyczna.
- Obszar krytyczny to taka część pola pod krzywą rozkładu, która stanowi obszar odrzucenia hipotezy zerowej (występujące w nim wartości statystyki testowej uprawniają do podjęcia takiej decyzji).
- Pozostała część to obszar, w którym wartości nie stanowią dostatecznej podstawy do odrzucenia hipotez zerowej

Obszar krytyczny

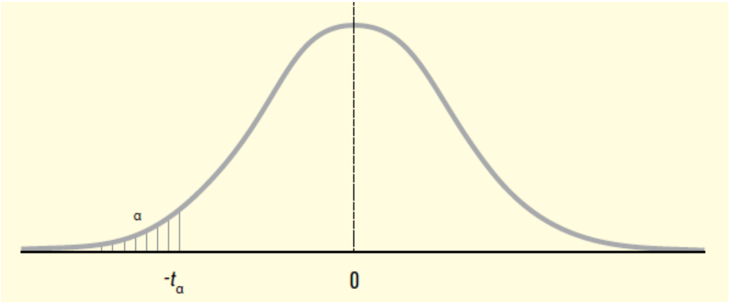
- Obszar krytyczny (a więc ten pozwalający na uznanie hipotezy zerowej za fałszywą) może być **jednostronny** lub **dwustronny** – tak jak test statystyczny





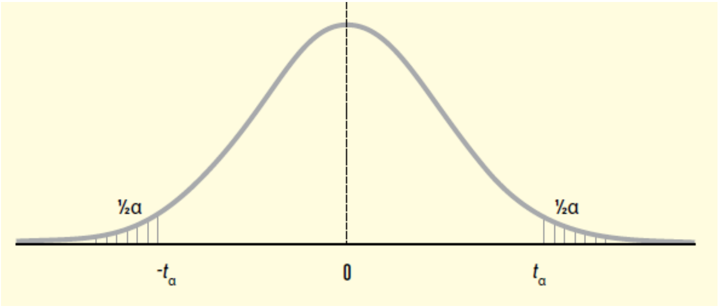
Jednostronny obszar krytyczny

Jeśli założony poziom istotności α wynosi 0,05 (czyli 5%), to obszar jednostronny wykorzystywany w testach jednostronnych stanowi 5% obszaru pola pod krzywą. Obszar ten może być lewostronny (jak na rysunku) lub prawostronny – w zależności od kierunku testowanej hipotezy statystycznej.



Dwustronny obszar krytyczny

- zakreskowane pola pojawiają się na dwóch krańcach rozkładu, a co tym idzie, pole 5% podzielone zostało na dwa pola o powierzchni 2,5% całego pola pod krzywą. Tego rodzaju obszary wykorzystywane są właśnie do testowania niekierunkowych hipotez statystycznych w testach dwustronnych.
- Statystyce testowej trudniej jest w związku z tym „wpaść” w ten obszar (testy dwustronne są mniej czułe, ich moc jest mniejsza), ale jeśli już „wpadnie”, to w zależności od znaku (wartość ujemna lub dodatnia) umożliwia wyciąganie wniosku o dowolnym kierunku efektu.



Obszar krytyczny

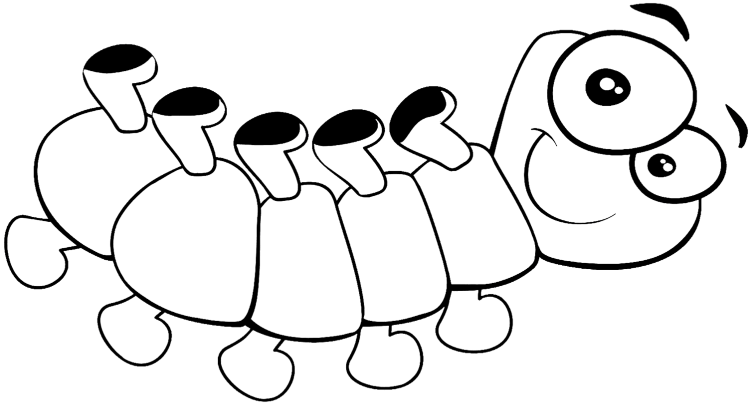
- Różne rozkłady prawdopodobieństwa mają inne wartości uznawane za krytyczne
- Dla rozkładu normalnego,
  - dla testu jednostronnego Z = +/-1,64
  - Dla testu dwustronnego Z = +/- 1,96

To właśnie dla tych wielkości uzyskane prawdopodobieństwo, przy założeniu, że hipoteza zerowa jest prawdziwa, wynosi p = 0,050 (czyli 5%,).

- Nie są one jednak stałe. Rozkład to niejedyny czynnik związany z wielkością wartości krytycznej. Innym ważnym elementem wyznaczania jej jest ustalenie **liczby stopni swobody**

Stopnie swobody

- Wartość p danego testu statystycznego wyliczana jest bowiem zawsze na podstawie wartości statystyki testowej o specyficznym rozkładzie (np. rozkładzie *t* Studenta), dla którego często należy określić właśnie ilości stopni swobody oraz kryterium istotności (najczęściej $\alpha = 0{,}05$).



Stopnie swobody

Definicja

- liczba stopni swobody (*df*, *degrees of freedom*) to liczba niezależnych wyników obserwacji pomniejszona o liczbę związków, które łączą te wyniki ze sobą
- liczbę tę można utożsamiać z liczbą niezależnych zmiennych losowych, które wpływają na wynik.
- liczba obserwacji minus liczba parametrów estymowanych za pomocą tych obserwacji.

Miara siły efektu

- Oprócz wartości *p* ważne jest jakie znaczenie ma ten wynik czyli jaką ma siłę efektu
- **Miara siły efektu** jest zwykle wystandaryzowaną (czyli niezależną od jednostki pomiaru) wartością wyliczoną na podstawie wyników otrzymanych w próbie, odpowiednią dla wybranego testu statystycznego, która świadczy o skali / wielkości badanego zjawiska (np. różnicy w średnich).
- Zawsze oprócz raportowania *p* powinno się zaraportować siłę efektu!

Dziękuję

Statystyki opisowe

Systematyczna Obserwacja

… czyli jak rozbić bank

Richard Jarecki

https://www.youtube.com/watch?v=dbr9lY1AezI

Statystyki opisowe

- Opisują zebrane przez Ciebie dane
- Nie pozwalają na wyciąganie jakichkolwiek wniosków nt. populacji
- Informują o danych w zebranej próbie
- **Jest to pierwszy krok od którego trzeba zacząć analizując dane**

Dane surowe

- ściśle określony zakres wartości (rozkład wylosowania poszczególnych numerów w ruletce nie mógłby przekroczyć zakresu 0–37),
- jeden lub kilka wierzchołków, opisujących lokalne maksimum lub maksima funkcji prawdopodobieństwa,
- szerokość rozkładu obrazująca rozproszenie obserwacji wokół wierzchołków.

Statystyki opisowe

Miary tendencji centralnej

Miary klasyczne:

- Średnia arytmetyczna
- Wszystkie pozostałe średnie np. geometryczna, ważona itp.

Miary pozycyjne

- Kwantyle
- Mediana
- Dominanta

Miary rozproszenia (rozrzutu, dyspersji)

- Rozstęp
- Wariancja i Odchylenie standardowe
- Rozstęp ćwiartkowy inaczej rozstęp międzykwartylowy IQR)

Miary tendencji centralnej

służą do opisania środka rozkładu – ukazania tego, co dzieje się w jego centrum.

MIARY KLASYCZNE

- są to miary oparte na wszystkich zebranych przez ciebie wynikach.
- każda obserwacja się liczy i ma wpływ na ostateczną wartość miary klasycznej
- ten rodzaj miar jest bardzo wrażliwy na obecność **przypadków odstających**.

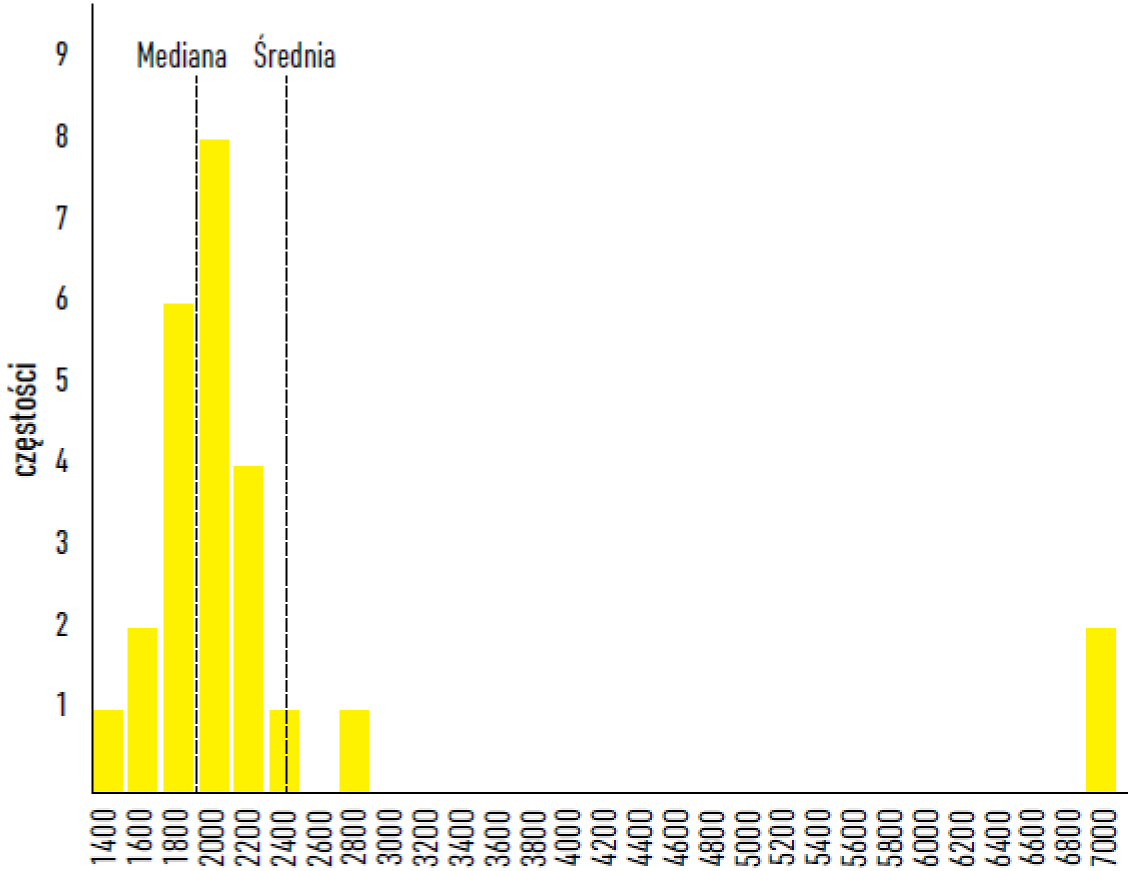
MIARY POZYCYJNE

- ich wynik określają jedynie niektóre z obserwacji znajdujących się w zebranym przez ciebie zbiorze danych.
- są niewrażliwe na występowanie obserwacji odstających.

Miary klasyczne

Średnia arytmetyczna

- Najprostsza miara tendencji centralnej
- Nie odporna na wartości odstające



Średnia trymowana

- Jest to sposób na radzenie sobie z odstającymi wynikami.
- Polega na odcięciu 5% lub 10 % skrajnych wyników

Miary pozycyjne

- Używane gdy spodziewasz się, że dane mogą być asymetryczne, występuje sporo obserwacji odstających lub pomiar jest na skali porządkowej

Kwantyle to określenie całej grupy statystyk obliczanych w podobny sposób, jednak różniących się nieco pewnym właściwościami. Dzielą zbiór wyników na określoną liczbę części, podając wartości, które stanowią w tym przypadku punkty podziału. W skład tej grupy wchodzi na przykład:

- **mediana (*Me* lub *Mdn*)**, czyli wartość środkowa dzieląca wyniki na dwie połowy,
- **kwartyle** zakładające podział wyników na 4 części
- czy mniej popularne **kwintyle** związane z podziałem na 5 części
- Czy też **percentyle** – dzielące na 100 części.

Percentyle

Przykład z matury

W części pisemnej/ustnej z …. uzyskał/a X%

Wynik taki sam lub niższy uzyskało Y% zdających

Percentyle

Przykład z matury

W części pisemnej/ustnej z …. uzyskał/a X%

Wynik taki sam lub niższy uzyskało Y% zdających

- z języka polskiego uzyskała wynik 79% (90 percentyl),
- z języka angielskiego 100% (100 percentyl),
- z matematyki 72% (98 percentyl).

Percentyle

Przykład z matury

W części pisemnej/ustnej z …. uzyskał/a X%

Wynik taki sam lub niższy uzyskało Y% zdających

- z języka polskiego uzyskała wynik 79% (90 percentyl),
- z języka angielskiego 100% (100 percentyl),
- z matematyki 72% (98 percentyl).

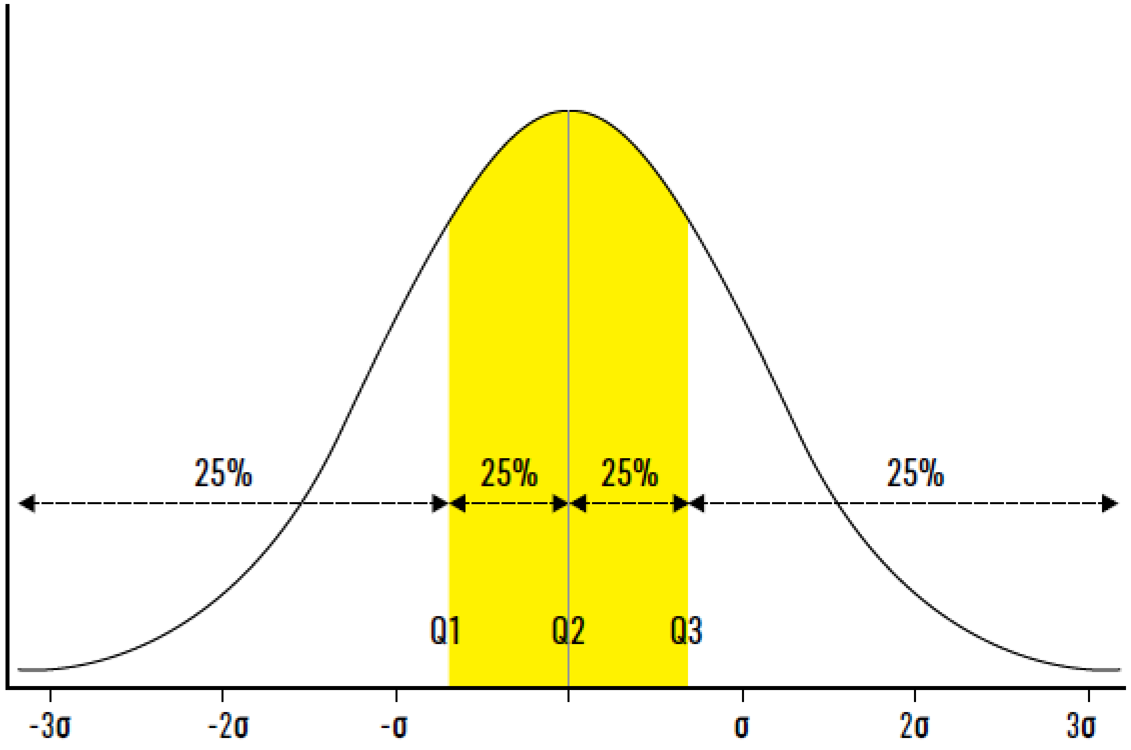
90 percentyl dla wyniku z języka polskiego oznacza, że 90% zdających napisało gorzej od niej lub na tym samym poziomie,

100 percentyl dla wyniku z języka angielskiego oznacza, że 100% zdających napisało gorzej od niej lub na tym samym poziomie,

98 percentyl z matematyki oznacza, że 98% zdających napisało gorzej od niej lub na tym samym poziomie.

Kwartyle

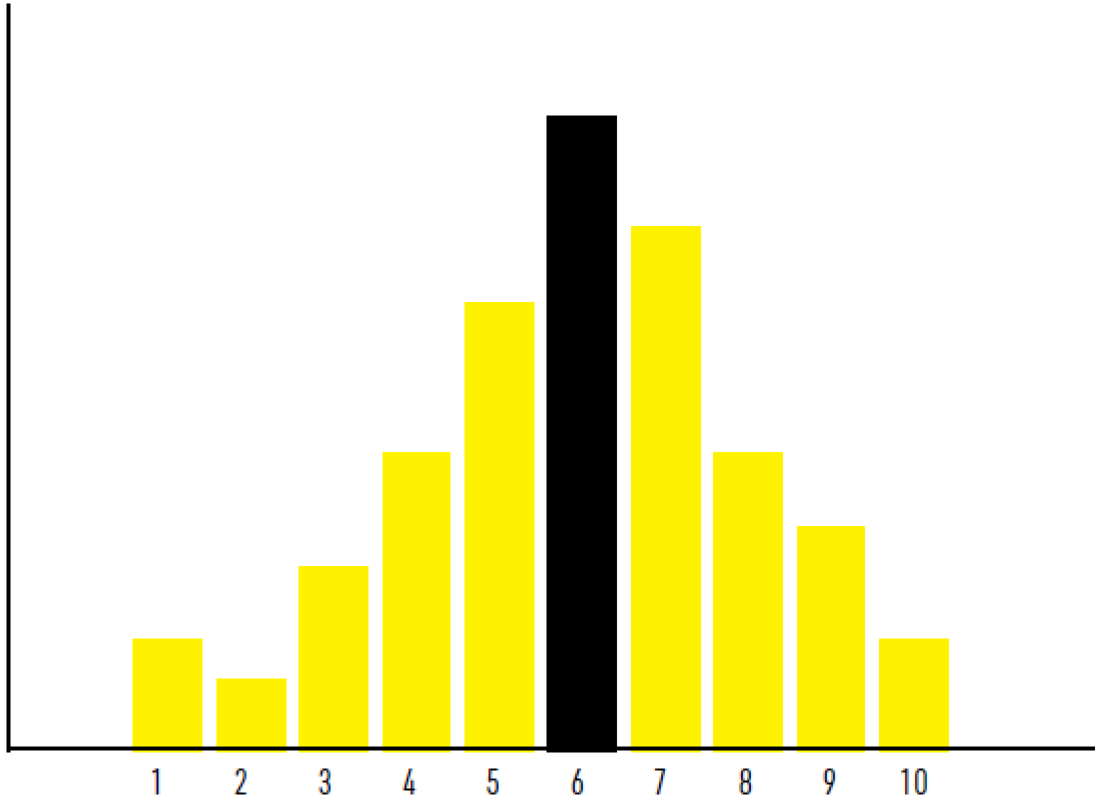
- Dzielą zbiór danych na cztery części
- **Kwartyl pierwszy (Q1)** to liczba wskazująca, że 25% twojego zbioru danych (inaczej jest to kwantyl 25 rzędu) ma wartości mniejsze bądź jej równe, a co za tym idzie 75% wartości jest od niej większych.
- **Kwartyl trzeci (Q3)** to liczba wskazująca, że 75% twojego zbioru danych (inaczej jest to kwantyl 75 rzędu) ma wartości mniejsze bądź jej równe, a co za tym idzie 25% wartości jest od niej większych.
- **Kwantyl drugi (Q2)** to inaczej **Mediana** – dzieli zbiór na równe połowy (inaczej kwantyl 50 rzędu)



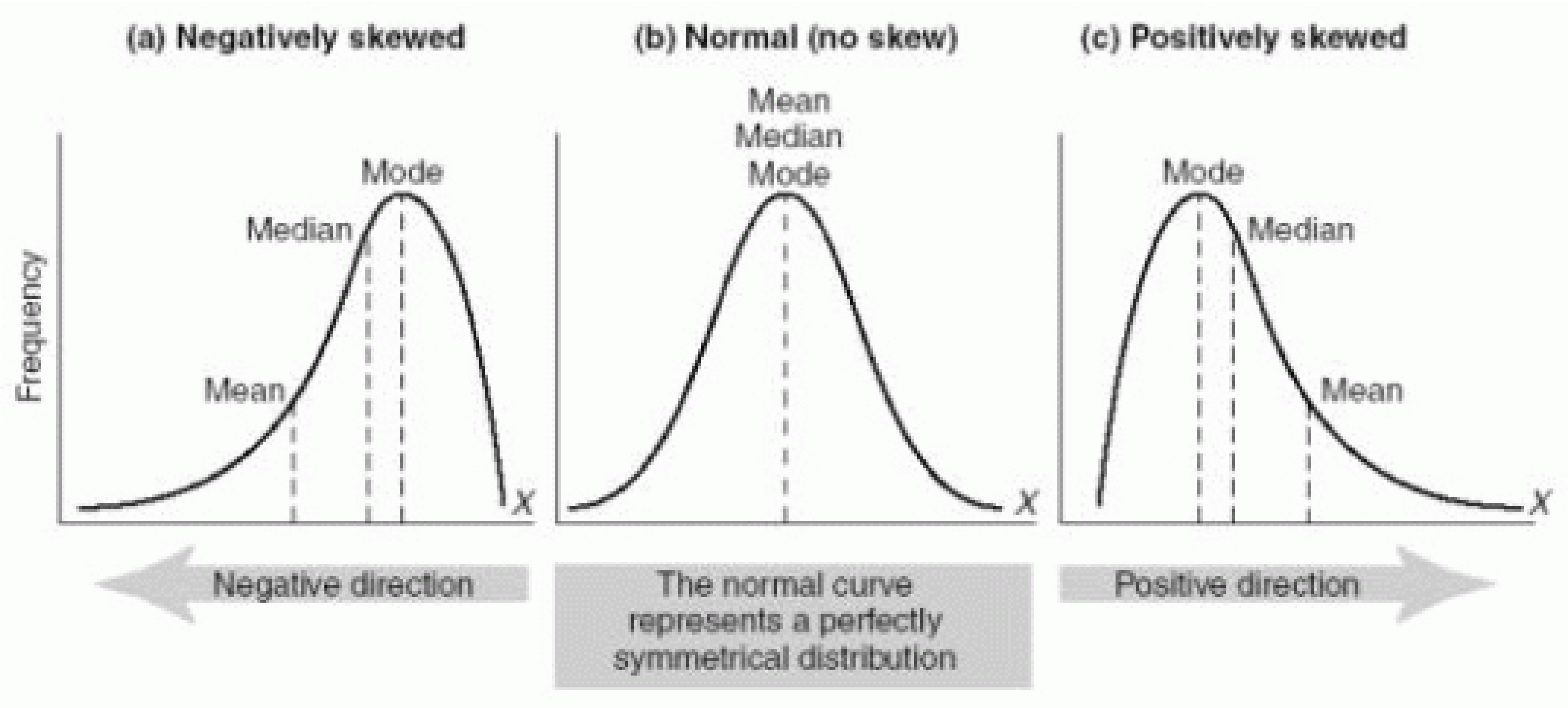
Miary pozycyjne

Dominanta, Moda, Modalna

- Mianem **dominanty (*D*)** (inaczej nazywanej też **modą** lub **modalną** – ***Mo***) określa się wartość, która występuje najczęściej w określonym zbiorze danych.
- używana miara w przypadku skali nominalnej oraz porządkowej



Trio: M, Me, D


(a) Negatively skewed
(b) Normal (no skew)
(c) Positively skewed
Frequency
Mode
Median
Mean
X
Mean
Median
Mode
X
Mode
Median
Mean
X
Negative direction
The normal curve represents a perfectly symmetrical distribution
Positive direction

Miary rozproszenia

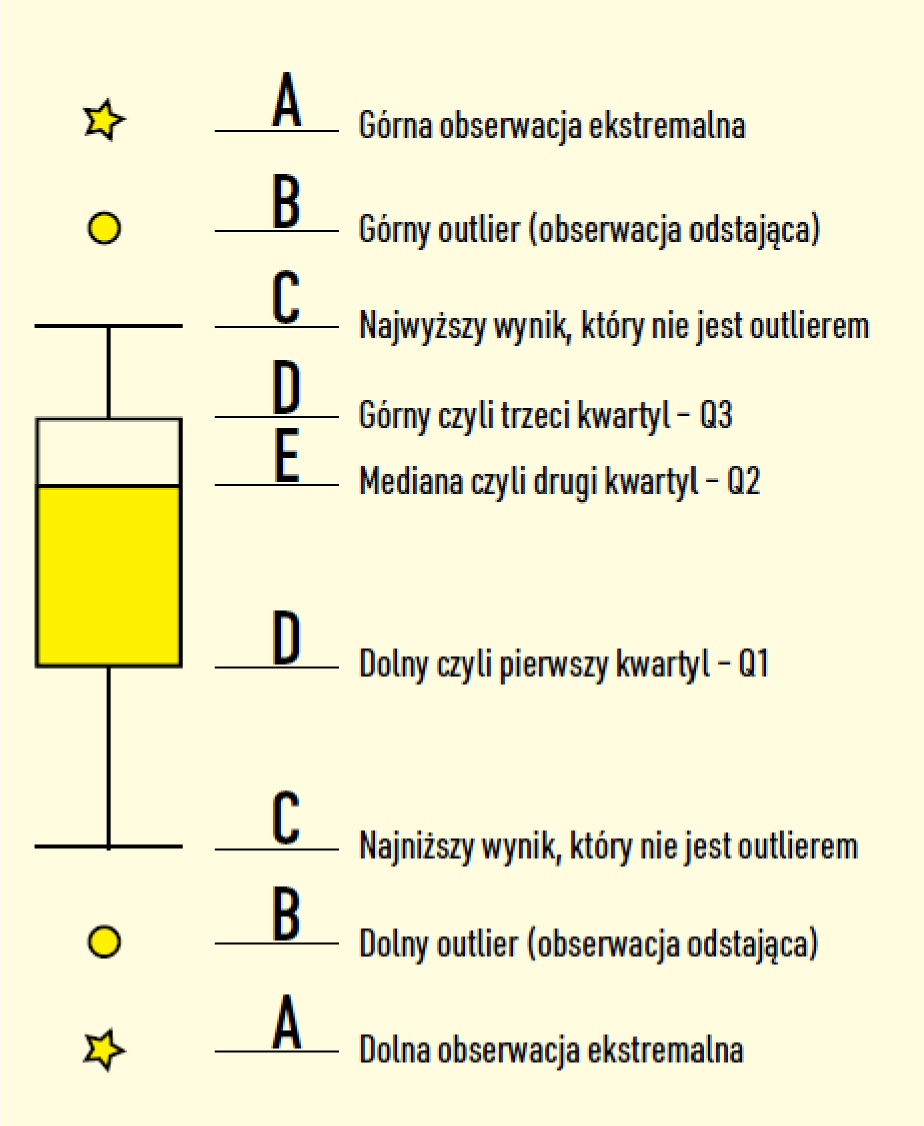
mówią one o pewnym rozłożeniu obserwacji względem tendencji centralnej, ale także o pewnej zmienności / zróżnicowaniu, jakie zachodzi w obrębie danych.

- **Rozstęp** (powiązany z wartością maksymalną i minimalną występującą w zbiorze danych),
- **wariancja i odchylenie standardowe** (jako klasyczne miary zmienności i odwołujące się rozproszenia wyników względem średniej arytmetycznej),
- • a także **rozstęp ćwiartkowy**, czy inaczej **rozstęp międzykwartylowy** – IQR (który nawiązuje do miar pozycyjnych, w tym przypadku do kwartyli).

Miary rozproszenia

Rozstęp

- **Rozstęp** oznacza różnicę pomiędzy **wartością maksymalną i minimalną** w zbiorze danych porządkowych lub ilościowych.
- rozstęp to miara, która mówi ci o tym, jaki jest maksymalny zakres wahań badanej przez ciebie cechy.
- Im większy, tym większych wartości pozostałych miar rozrzutu możesz się spodziewać.
- Nienaturalnie duża wartość rozstępu może również wskazywać na obecność wartości odstających.

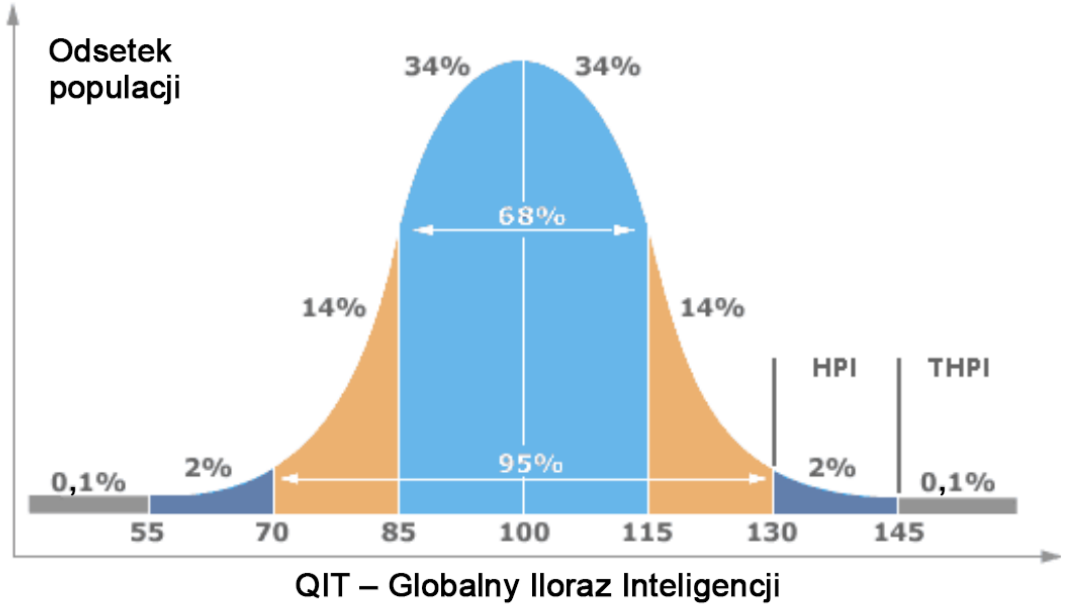


Miary rozproszenia

Wariancja i odchylenie standardowe

- **Wariancja (*var*, s2, σ2)** w bardzo prosty sposób pokazuje stopień odchylenia danych od średniej arytmetycznej.
- Im większa jest wartość wariancji, tym bardziej dane są rozproszone, a więc mniej obserwacji znajduje się w centrum rozkładu, a więcej na jego końcach (ogonach).
- Z kolei im mniejsza jest wariancja, tym bardziej wartości badanej przez ciebie cechy są skupione wokół średniej.
- mała wartość wariancji wskazuje, że średnia arytmetyczna jest dobrym przybliżeniem wartości cechy, którą badasz. Możesz powiedzieć, że średnia bardzo dobrze oddaje tendencję centralną i możesz ją stosować jako rzetelną charakterystykę przeciętnego poziomu twojej zmiennej w badanej próbie.
- Duże wartości wariancji wskazują natomiast, że średnia niedokładnie odzwierciedla rozkład danej cechy. Mamy bowiem całkiemsporo obserwacji nad i pod średnią, które są od niej znacznie oddalone.

- **odchylenie standardowe (SD, s, σ)** jest pierwiastkiem kwadratowym z wariancji



Source : http://www.cabinetorientation.fr/

**HPI: Wysoki Potencjał Intelektualny

***THPI: Bardzo Wysoki Potencjał Intelektualny

Wariancja

to suma kwadratów różnic pomiędzy wartościami obserwacji a średnią wyliczoną dla całego zbioru danych podzielona przez liczbę obserwacji (lub liczbę obserwacji minus jeden).

PRZYKŁAD

Wyobraź sobie, że zbiór twoich danych wygląda następująco:

2	2	2

M = 2

- *Krok 1*

Policz różnice między każdą obserwacją a średnią (definicja: ...*różnic pomiędzy wartościami obserwacji a średnią wyliczoną dla całego zbioru danych*…).

Wynikami tych działań będą kolejno 0, 0 i 0?

Wariancja

- *Krok 2*

Policz kwadraty wyżej wymienionych różnic! (definicja: …*kwadratów różnic*…). Innymi słowy podnieś każdy z wyników (w tym przypadku każde 0) do kwadratu.

wynikami tych działań nadal będą kolejno 0, 0, i 0

- *Krok 3*

Dodaj do siebie uzyskane kwadraty różnic (definicja: *Suma kwadratów różnic*…)

wynik = 0!

- *Krok 4*

Podziel wynik przez liczbę obserwacji (lub liczbę obserwacji minus 1)! (definicja: *Suma… podzielona przez liczbę obserwacji lub liczbę obserwacji minus jeden*).

Liczba obserwacji: 3 (bo 2, 2, 2).

Liczba obserwacji minus 1 = 2.

W obydwu przypadkach podzielenie 0 przez powyższe liczby i tak da wynik 0, a zatem Wariancja = 0!

Wariancja przykład 2

przez N - 1, gdy mamy wyniki próby,
przez samo N, gdy mamy wyniki populacji.

- Twój zbiór cyfr wynosi: 2, 3, 4

Krok 1 średnia

- M = 3

Krok 2

Policz kwadraty wyżej wymienionych różnic! (definicja: …*kwadratów różnic*…). Innymi słowy podnieś każdy z wyników (do kwadratu.

1,0,1

Krok 3

Dodaj do siebie uzyskane kwadraty różnic

1 + 0 + 1 = 2

Krok 4

Podziel wynik przez liczbę obserwacji (lub liczbę obserwacji minus 1)! (definicja: *Suma… podzielona przez liczbę obserwacji lub liczbę obserwacji minus jeden*).

Liczba obserwacji: 3 (bo 2,3,4).

Liczba obserwacji minus 1 = 2.

Wariancja = 2 (suma)/3 = 0,66 lub 2/2=1

Odchylenie standardowe

To pierwiastek z wariancji

W 1 przykładzie $\sqrt{0} = 0$

W 2 przykładzie $\sqrt{1} = 1$

Klasyczny współczynnik zmienności V

- Pozwala na porównanie różnych zmienności między sobą i sprawdzenie jak ona jest duża
- wzór sprowadza się do podzielenia wartości odchylenia standardowego przez średnią uzyskaną w próbie i pomnożenie wyniku przez 100.

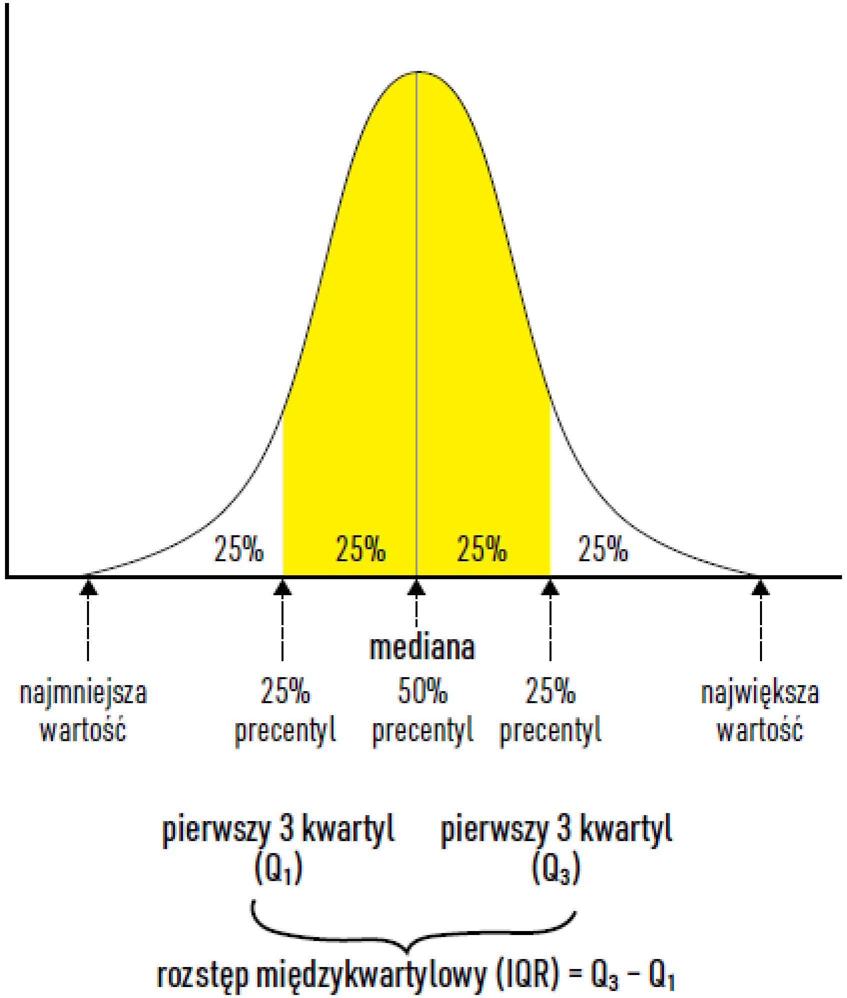
- W 1 wypadku
- 0/0*100=0
- W 2 przykładzie
- 1/3*100 = 33,33

Wartość współczynnika zmienności	Interpretacja
<25%	Niewielka zmienność
25% – 50%	Umiarkowana zmienność
50% – 100%	Duża zmienność
>100%	Bardzo duża zmienność

Miary pozycyjne

IQR – Rozstęp międzykwartylowy

- jest jako różnica pomiędzy trzecim a pierwszym kwartylem.
- mówi bowiem o szerokości przedziału zawierającego 50% środkowych wartości danego rozkładu prawdopodobieństwa
- Im jest on szerszy, tym większe jest rozproszenie wyników wokół centrum rozkładu
- jest on mierzony w tych samych jednostkach, co badana cecha, co zdecydowanie ułatwia jego interpretację.

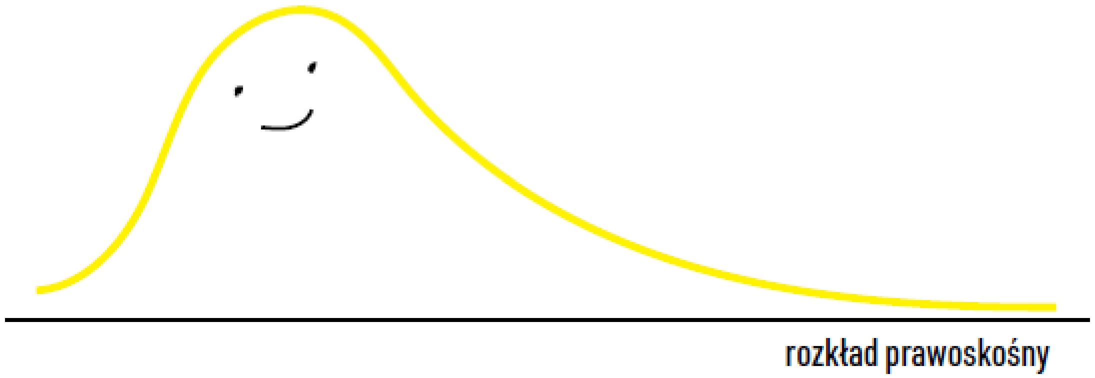


Miary asymetrii i obserwacji odstających

Skośność

- **Skośność** jest miarą określającą asymetrię rozkładu danych.
- Informuje przy tym zarówno o kierunku tej asymetrii, jak i o jej mocy. Jest bardzo użyteczna w kontekście sprawdzania normalności rozkładu i adekwatności danych do użycia testów parametrycznych.
- Wartość centralną dla współczynnika skośności stanowi 0 (idealnie symetryczny rozkład normalny taką właśnie wartością skośności się odznacza).

- **rozkład prawoskośny** – odznacza się pozytywnymi (> 0) wartościami współczynnika skośności. Jego prawe ramię na wykresie gęstości jest dłuższe niż lewe
- **rozkład lewoskośny** – odznacza się negatywnymi (< 0) wartościami współczynnika skośności. Jego lewe ramię na wykresie gęstości jest dłuższe niż prawe
- **Rozkład symetryczny** – wartość współczynnika skośności wynosi 0, co oznacza że obydwa ramiona rozkładu mają taką samą długość

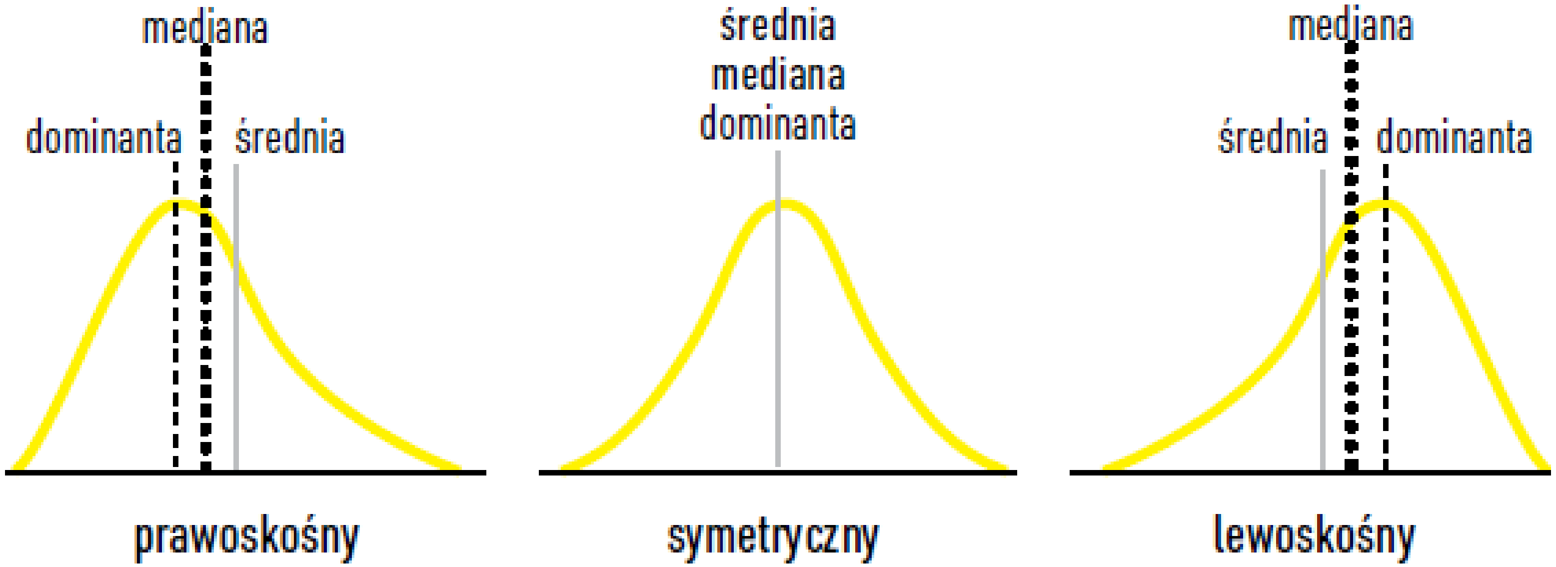


Skośność cd.

- Wartość skośności stosuje się często przy ocenie przydatności danych do użycia testów parametrycznych.
- Testy parametryczne bardzo cenią sobie symetryczne rozkłady o niskiej wartości skośności, ponieważ wtedy ich oszacowania są najdokładniejsze.
- Wykonując test parametryczny na danych **mocno** skośnych, narażasz się na zwiększony błąd pomiaru i niewłaściwie wyznaczony poziom istotności.
- Zazwyczaj przyjmuje się, że dla rozkładów mających wartości skośności znajdujące się w przedziale od **–2 do 2** można stosować testy parametryczne.
- Powyżej tej granicy stosuje się testy nieparametryczne

Skośność a miary tendencji centralnej

- dla rozkładu symetrycznego wartość średniej arytmetycznej, mediany i mody jest taka sama,
- **dla rozkładu lewoskośnego** najmniejszą wartość przyjmuje średnia arytmetyczna, pośrodku znajduje się mediana, a ostatnia jest moda, czyli wartość najczęstsza.
- **dla rozkładu prawoskośnego** wartość mody jest najmniejsza z wszystkich parametrów tendencji centralnej; na następnym miejscu plasuje się mediana, a ostatnia jest średnia arytmetyczna,



Kurtoza

Zwana **współczynnikiem ekscesu**

- Kurtoza informuje cię po pierwsze o tym, jak wyglądają ogony rozkładu prawdopodobieństwa twojej zmiennej, a po drugie o obecności obserwacji odstających w twoich danych.
- Na jej podstawie możesz powiedzieć zatem, czy badana zmienna charakteryzuje się wartościami, które znacząco odbiegają od pozostałych w kierunku dodatnim lub ujemnym.
- Centralną wartością kurtozy, podobnie jak w przypadku współczynnika skośności jest 0
- Nie ma górnej i dolnej granicy określającej minimalne lub maksymalne wartości tej miary.

Kurtoza

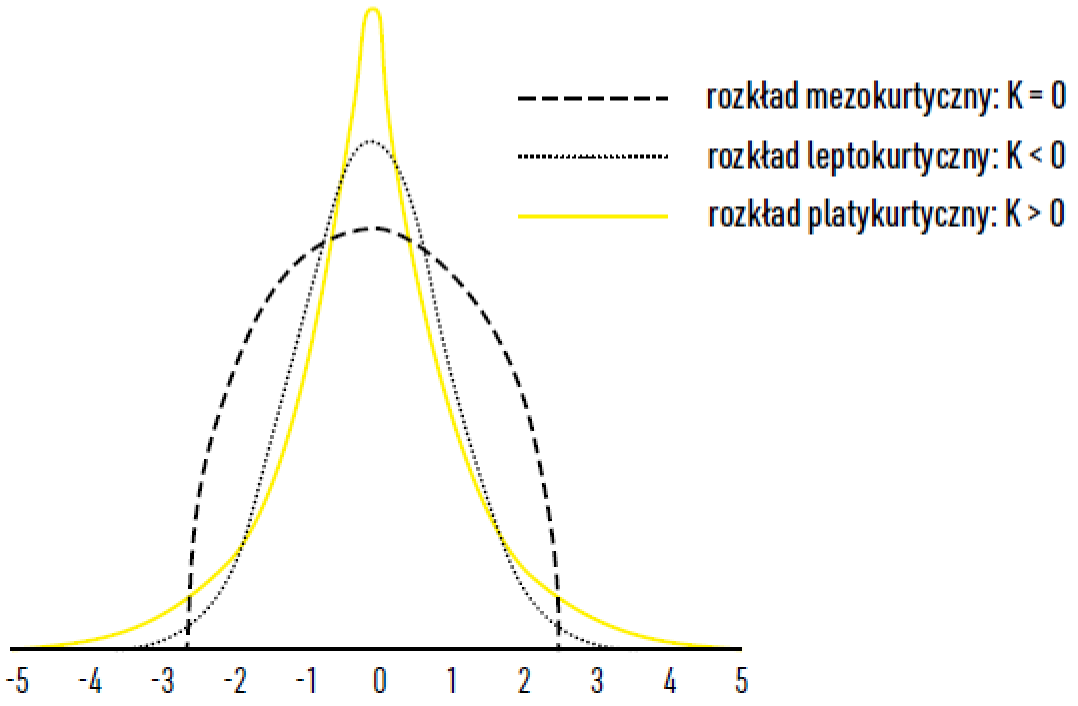
- **kurtoza dodatnia:** oznacza to, że ogony rozkładu prawdopodobieństwa badanej przez ciebie cechy są większe niż ogony rozkładu normalnego.
- prawdopodobieństwo krańcowych wartości jest znacznie większe niż mógłbyś się tego spodziewać na podstawie rozkładu normalnego;
- jeżeli wartość kurtozy jest odpowiednio duża (powyżej 2, a czasem nawet już od 1,5), powinieneś przeszukać swoje dane pod kątem obecności obserwacji odstających.

- **kurtoza ujemna:** oznacza to, że ogony rozkładu prawdopodobieństwa badanej przez ciebie cechy są znacznie węższe niż ogony rozkładu normalnego lub niema ich wcale.
- Negatywna wartość kurtozy w zasadzie w zupełności wyklucza możliwość istnienia obserwacji odstających w badanym przez ciebie rozkładzie.

Kurtoza

- Rozkład charakteryzujący się <u>dodatnią</u> wartością kurtozy określany jest jako **platykurtyczny**,
- Rozkład charakteryzujący się <u>ujemną</u> wartością kurtozy określany jest jako **leptokurtyczny**,

- kurtoza = 0: oznacza to, że prawdopodobieństwo wystąpienia obserwacji odstających jest takie samo jak w rozkładzie normalnym.
- Rozkład charakteryzujący się zerową wartością kurtozy określany jest jako **mezokurtyczny**.



Testy normalności rozkładu

- stosuje się je przy okazji obliczania współczynników skośności i kurtozy
- Ich **hipoteza zerowa** zakłada zgodność rozkładu danych z rozkładem normalnym, natomiast
- **hipoteza alternatywna** wskazuje, że istnieją pewne różnice

- **test Shapiro-Wilka (test S-W):** może być stosowany do sprawdzenia zgodności rozkładu danych z rozkładem normalnym w przypadku prób o wielkości od 3 do 5000 obserwacji.
- Jego stosowanie bezwzględnie zaleca się w przypadku prób mniejszych niż 50 obserwacji. Wykazuje się wtedy znacznie większą mocą niż pozostałe testy,
- **test Kołmogorowa-Smirnowa (test K-S):** sprawdza dopasowanie danych do dowolnego rozkładu o znanych parametrach (w przypadku rozkładu normalnego są to średnia i odchylenie standardowe). Wykazuje się zazwyczaj znacząco niższą mocą, niż test S-W.
- Z tego też powodu stosuje się go jedynie przy wystarczająco dużych liczebnościach próby (zwykle chodzi o $n > 2000$)

Testy normalności rozkładu

- UWAGA! **w przypadku dużej liczebności badanej próby zdecydowanie lepszym rozwiązaniem niż testowanie normalności rozkładu jest zwyczajne obliczenie współczynników skośności i kurtozy**.
- Wraz ze wzrostem liczebności testy normalności rozkładu mają tendencję do wskazywania wyników istotnych nawet przy bardzo niewielkich odchyleniach od normalności.
- A zatem, **jeśli twoja próba obejmuje więcej niż choćby 100 osób (a już z pewnością wtedy, gdy są to próby o wielkości kilkuset lub kilku tysięcy osób), oprócz wyników ww. testów, przyjrzyj się samym histogramom, a także miarom asymetrii.**
- **Zastosuj wówczas zasadę przedziału -2 do 2 dla skośności -1,5 do +1,5 dla kurtozy.**
- **Jeśli wartości skośności i kurtozy mieszczą się w tym zakresie, możesz przyjąć, że analizowany rozkład, choć nie jest do końca normalny, to odstępstwo to nie powinno spowodować większych problemów.**

Dziękuję

KORELACJA

Natalia Józefacka

KORELACJA

- Analiza korelacji parametrycznej, nazywana inaczej korelacją *r* Pearsona służy do badania związków.
- Wykorzystując **współczynnik korelacji**, można sprawdzić **siłę** i **kierunek** relacji między zmiennymi.
- Przyjmuje wartości od -1 do +1, a im bliżej |1|, tym silniejszy jest związek (0 oznacza natomiast brak związku!).
- dodatni lub ujemny znak tego wskaźnika informuje o kierunku zależności (mówi się wówczas o związku **pozytywnym / dodatnim** lub **negatywnym / ujemnym**).

MECHANIZM TESTU

wskaźnik korelacji jest tak naprawdę wystandaryzowanym współczynnikiem kowariancji.

Sama kowariancja wskazuje na kierunek relacji, ale po wystandaryzowaniu (do postaci *r* Pearsona) można ocenić również jej siłę.

Na samym końcu należy stwierdzić, czy obliczona relacja jest istotna i tu z pomocą przychodzi test *t* Studenta.

322,4 / 9 = 35,82

MECHANIZM TESTU KOWARIANCJA

Aby obliczyć współczynnik korelacji, należy obliczyć **kowariancję**

	KROK 1		KROK 2		KROK 3
KOLUMNA	A	B	C	D	E
	Zmienna X	Zmienna Y	Średnia X (9,7) minus wynik osoby X	Średnia Y (26,8) minus wynik osoby Y	Wynik mnożenia (C*D)
Ob 1	10	27	-0,3	-0,2	0,06
Ob 2	16	40	-6,3	-13,2	83,16
Ob 3	13	37	-3,3	-10,2	33,66
Ob 4	9	14	0,7	12,8	8,96
Ob 5	16	30	-6,3	-3,2	20,16
Ob 6	5	21	4,7	5,8	27,26
Ob 7	2	11	7,7	15,8	121,66
Ob 8	7	17	2,7	9,8	26,46
Ob 9	9	31	0,7	-4,2	-2,94
Ob 10	10	40	-0,3	-13,2	3,96
średnia	9,7	26,8		SUMA	322,4
SD	4,47	10,64			KROK4

SD – odchylenie standardowe

KORELACJA

Współczynnik korelacji jest obliczany podobnie jak kowariancja z jedną różnicą – w pierwszym kroku wszystkie wartości surowe zamieniane są na skalę *Z*.

Po standaryzacji kolejne kroki są identyczne jak w przypadku obliczania wskaźnika kowariancji.

Na szczęście wszystkie te elementy wyliczy program statystyczny!

Sam test mówiący o istotności statystycznej współczynnika opiera się na rozkładzie *t* Studenta

Skonstruowany w ten sposób test, odwołując się do liczebności badanej grupy, powoduje, że przy dostatecznej jej wielkości nawet niskie (a co za tym idzie świadczące o słabych związkach) współczynniki korelacji mogą okazać się w analizie istotne.

HIPOTEZY

współczynnik korelacji jest opisowym testem statystycznym,

wnioskowanie na jego podstawie polega na określeniu, czy wyliczona wartość jest istotna statystycznie, a konkretnie, czy jest istotnie różna od zera.

hipoteza zerowa (H0) głosi, że prawdziwa wartość *r* (a więc współczynnik korelacji w populacji) wynosi 0 (tj. nie ma związku);

hipoteza alternatywna (H1) brzmi natomiast, że $r \neq 0$.

PRZYKŁADY HIPOTEZ

Hipotezy statystyczne

(+)

H_0 (zerowa): *Liczba myśli o seksie w miesiącu nie jest związana z liczbą przeczytanych w tym okresie książek.*

H_1 (alternatywna bezkierunkowa): *Liczba myśli o seksie w miesiącu jest związana z liczbą przeczytanych w tym okresie książek.*

(-)

H_0 (zerowa): *Liczba myśli o seksie w miesiącu nie jest związana z liczbą obejrzanych w tym okresie seriali.*

H_1 (alternatywna bezkierunkowa): *Liczba myśli o seksie w miesiącu jest związana z liczbą obejrzanych w tym okresie seriali.*

ZAŁOŻENIA

Założenia:

• Dwie zmienne na **skali ilościowej**

• Dane są **sparowane** – potrzebujesz wyników **obu zmiennych dla każdej obserwacji.**

• Obserwacje **niezależne** – każda obserwacja to **inny obiekt, osoba, zjawisko.**

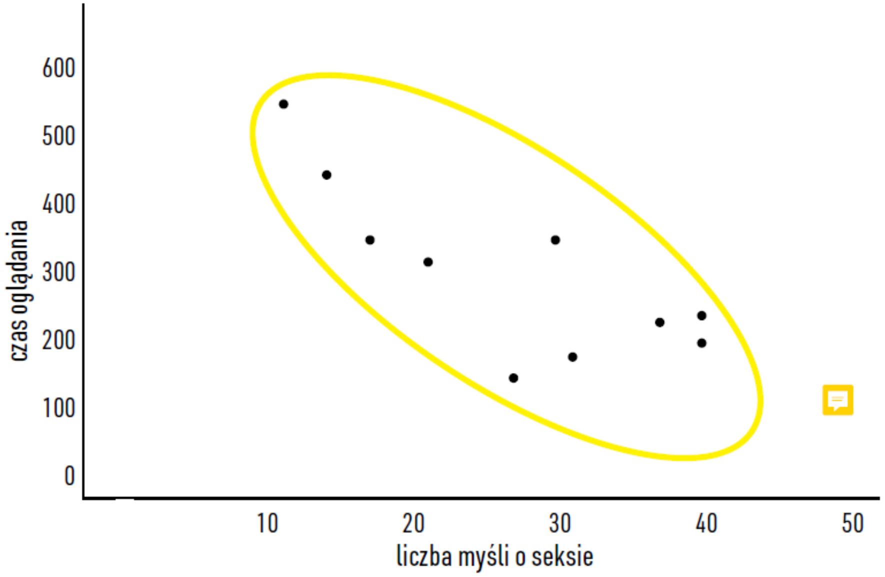
Diagnostyka rozkładu:

• **Prostoliniowość** – między zmiennymi – dwie zmienne są ze sobą powiązane w taki sposób, że **zmianie wartości jednej zmiennej towarzyszą przewidywalne zmiany drugiej zmiennej**

• Wartości **odstające** – uważaj szczególnie na te **wpływowe**!

• Rozkład **normalny** – dotyczy rozkładu **próbkowania (UWAGA! Jeżeli liczba obserwacji przekracza 30 os., nie musisz bezwzględnie sprawdzać tego założenia).**

WYKRES ROZRZUTU



Wykorzystywany do:

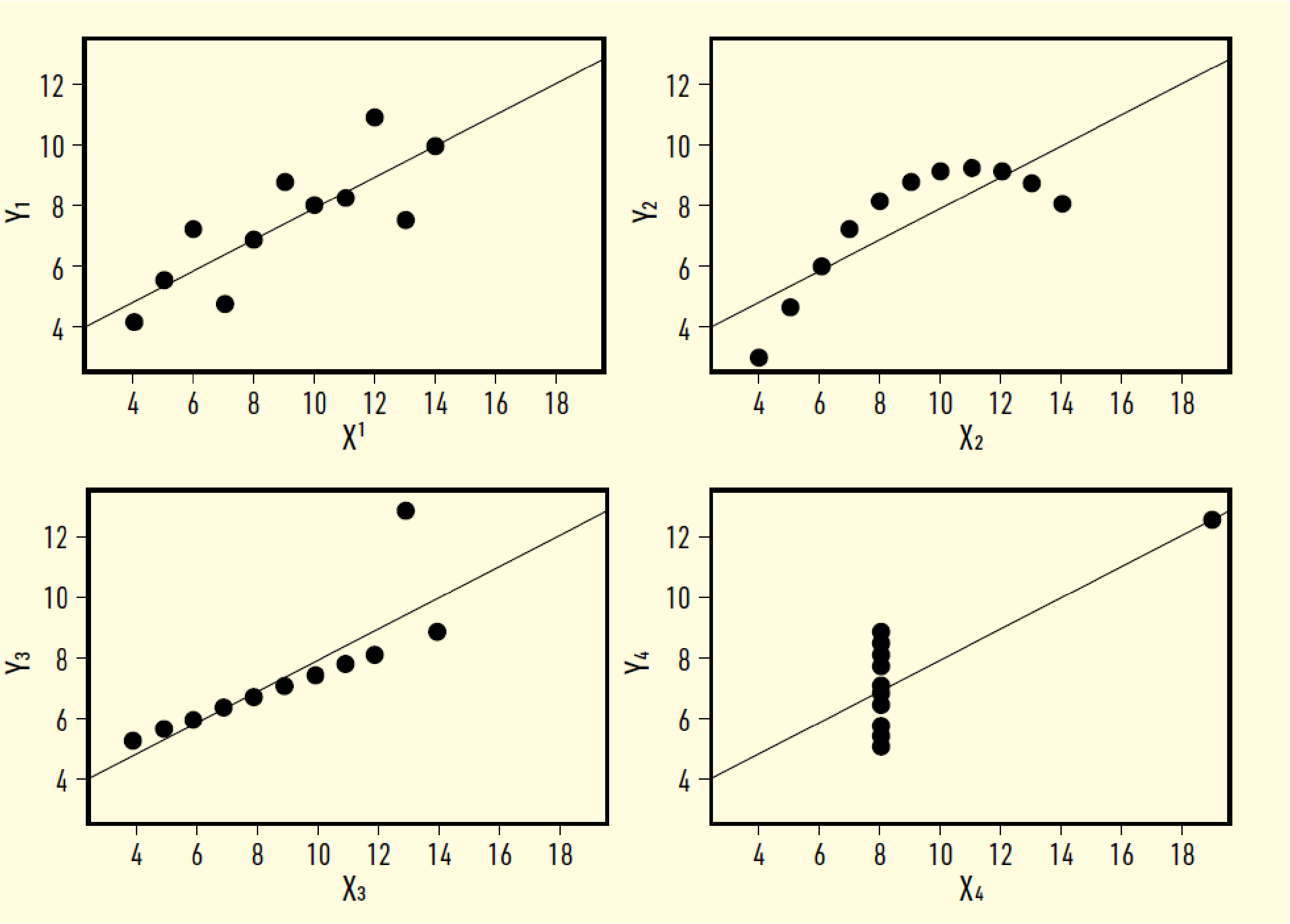
- Diagnostyki liniowości i obserwacji odstających
- Można zaobserwować kierunek zależności
- Im węższa elipsa tym silniejszy związek

KWARTET ANSCOMBE’A

- Średnia zmiennej x = 9 Wariancja zmiennej x = 10

• Średnia zmiennej y = 7,5 Wariancja zmiennej y = 3,75

• Współczynnik korelacji r = 0,816



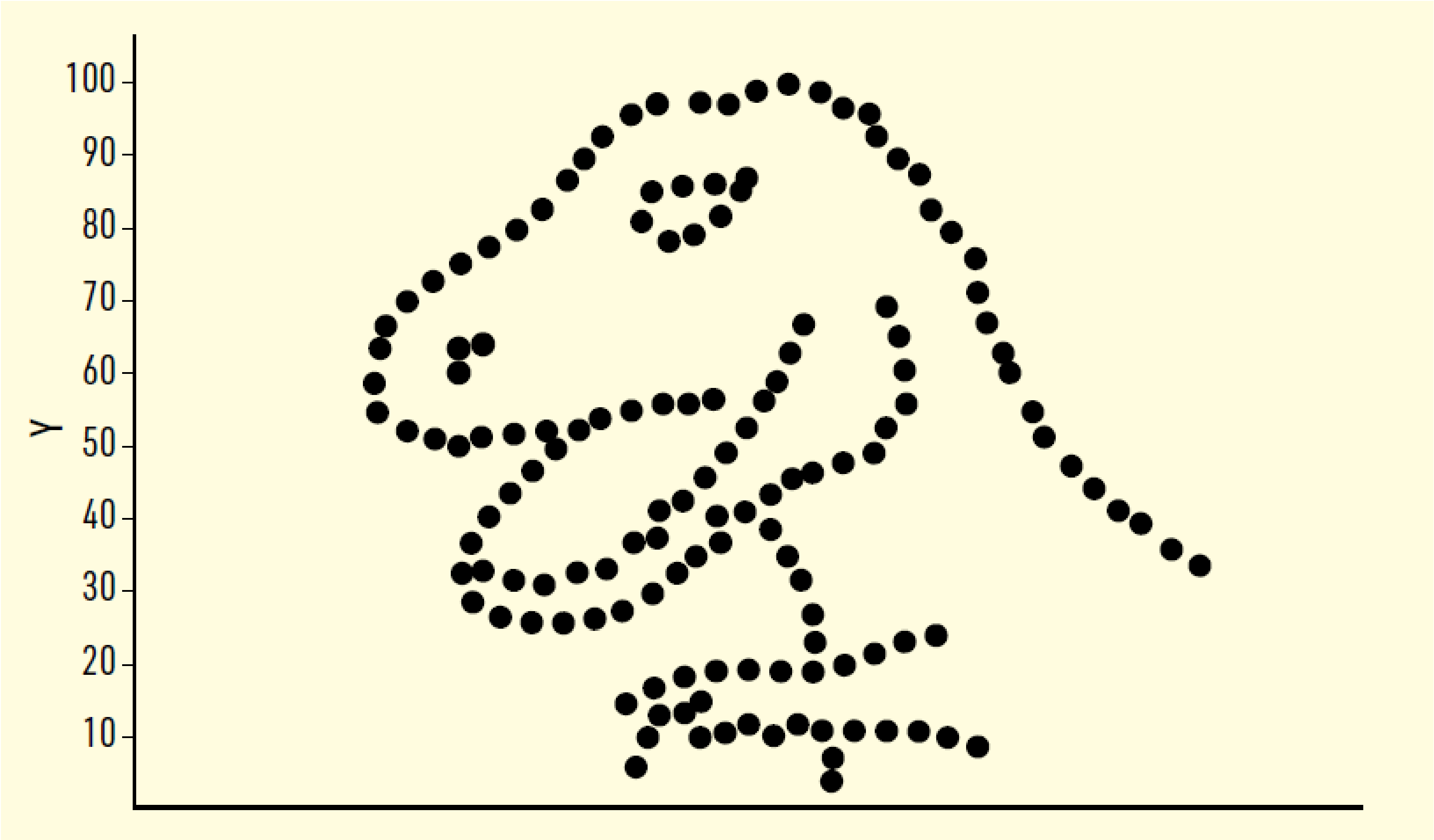
http://robertgrantstats.co.uk/drawmydata.html

ROBERT GRANT

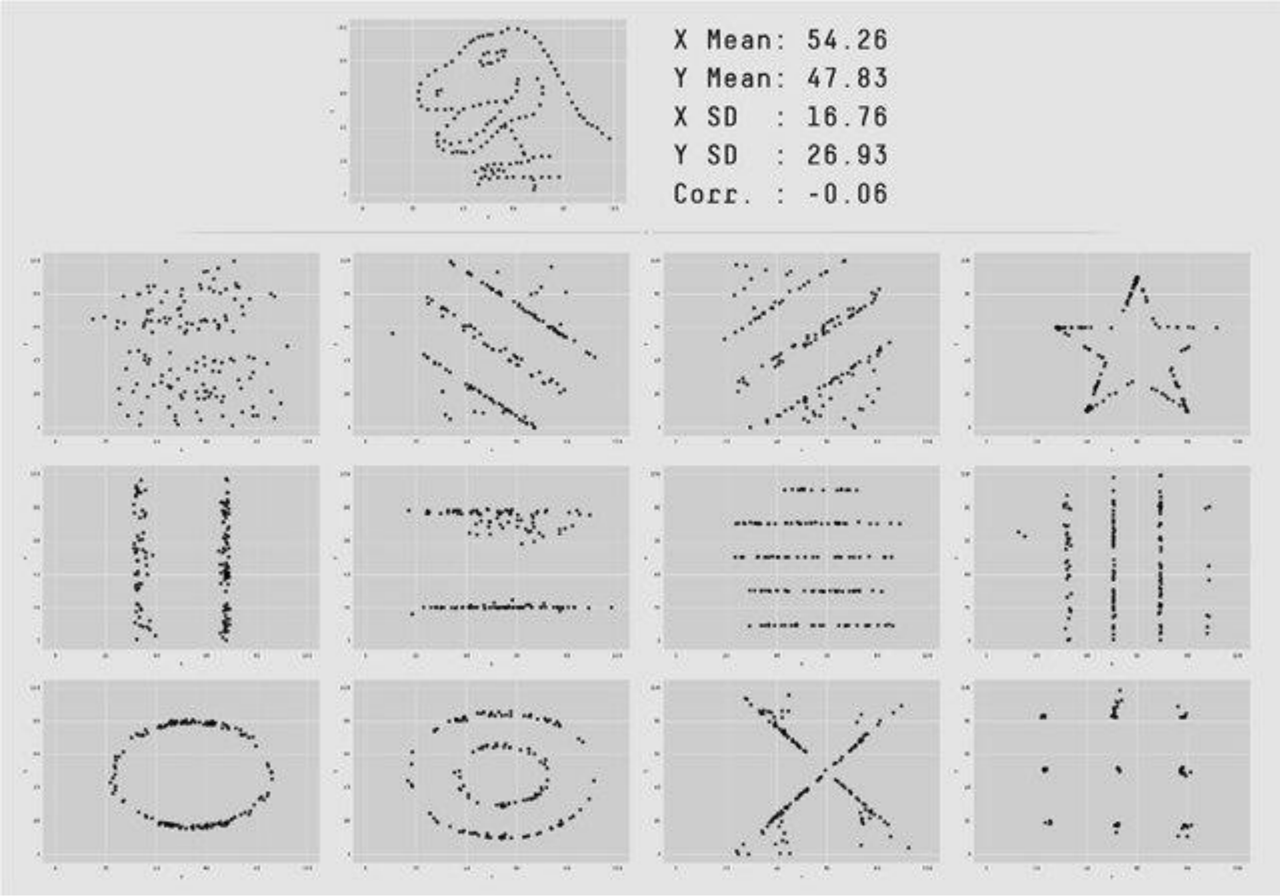
Narzędzie to zostało wykorzystane w stworzeniu słynnego Datasaurusa powstałego

na podstawie 142 obserwacji, dla których wyliczony współczynnik korelacji

wynosił -0,06.



MATEJKA I FITZMAURIC (2017)



r **Pearsona**

INTERPRETACJA

Tabela 8.3. Interpretacja siły korelacji (Cohen, 1988)

Siła korelacji	Interpretacja
0,1–0,2	Słaba zależność
0,3–0,5	Umiarkowana zależność
> 0,6	Silna zależność

Tabela 8.4. Interpretacja siły korelacji (Gignac, Szodorai, 2016)

Siła korelacji	Interpretacja
0,1	Słaba zależność
0,2	Umiarkowana zależność
0,3 i wyżej	Silna zależność

Źródło: Gignac G. E., Szodorai E. T. (2016). Effect size guidelines for individual differences researchers. Personality and Individual Differences, 102, 74–78; doi:10.1016/j.paid.2016.06.069

INTERPRETACJA

współczynnik determinacji (R2) jest definiowane jako część zmienności jednej ze zmiennych w korelacji, którą tłumaczy się zmiennością drugiej zmiennej

Dla analizowanego związku (hipotezy) $R^2 = 0{,}58$ (tyle wynosi podniesienie do kwadratu wartości $r = 0{,}76$!).

Możesz to zinterpretować w taki sposób, że za pomocą liczby przeczytanych książek można wyjaśnić zróżnicowanie wyników liczby myśli o seksie w 58%. To znaczy, że 42% zmienności wyjaśniane jest przez inne czynniki.

PODSUMOWANIE

Podsumowując, **dzięki analizie współczynnika korelacji można się dowiedzieć o sile i kierunku związku. Podnosząc współczynnik korelacji do kwadratu, otrzymuje się współczynnik determinacji, który informuje, ile zmienności wyników zmiennej Y można wyjaśnić przez zmienną X.**

PRZYCZYNOWOŚĆ

Określenie przyczynowości jest możliwe przy spełnieniu trzech warunków jednocześnie:

• **Dwie zmienne są ze sobą powiązane** – to sprawdzasz wykonując analizę korelacji i w przypadku powyżej badanej relacji uzyskałeś potwierdzenie!

• **Pierwszeństwo czasowe** – jedna zmienna musi pojawić się wcześniej, tak jak przyczyna zawsze występuje przed skutkiem. W kontekście przeczytanych książek i posiadanych myśli trudno powiedzieć, co pojawiło się jako pierwsze.

• **Kontrola trzeciej zmiennej** – czasami nazywanej **zmienną zakłócającą**. Dla ustalenia zależności przyczynowo-skutkowej należy udowodnić, że zmienna X powoduje zmiany w zakresie Y i **tego związku nie można wyjaśnić w inny sposób**, gdyż wszystkie pozostałe zmienne są kontrolowane. Jak widzisz, tego kryterium badania korelacyjne zwykle nie uwzględniają https://www.tylervigen.com/spurious-correlations

3 ZMIENNA

Problem **trzeciej zmiennej** występuje, gdy związek między analizowanymi zmiennymi jest tak naprawdę pośredni, a odpowiada za to inny, nieanalizowany w badaniu czynnik.

Problem trzeciej zmiennej wiąże się m.in. ze zjawiskiem tzw. **korelacji pozornych**

WIELKOŚĆ PRÓBY A POZIOM WSPÓŁCZYNNIKA KORELACJI

r	N	r	N	r	N	r	N
0,01	38146	0,26	58	0,51	16	0,76	7
0,02	9605	0,27	54	0,52	15	0,77	7
0,03	4269	0,28	50	0,53	15	0,78	7
0,04	2402	0,29	47	0,54	14	0,79	7
0,05	1538	0,3	44	0,55	14	0,80	7
0,06	1068	0,31	41	0,56	13	0,81	7
0,07	785	0,32	39	0,57	13	0,82	6
0,08	601	0,33	36	0,58	12	0,83	6
0,09	475	0,34	34	0,59	12	0,84	6
0,10	385	0,35	32	0,6	12	0,85	6
0,11	319	0,36	31	0,61	11	0,86	6
0,12	268	0,37	29	0,62	11	0,87	6
0,13	228	0,38	28	0,63	11	0,88	5

Peternek, P., & Kośny, M. (2011). Kilka uwag o testowaniu istotności współczynnika korelacji. *Zesz. Nauk. WSB Wroc*, *20*, 341–350.

DZIĘKUJĘ

Test t Studenta

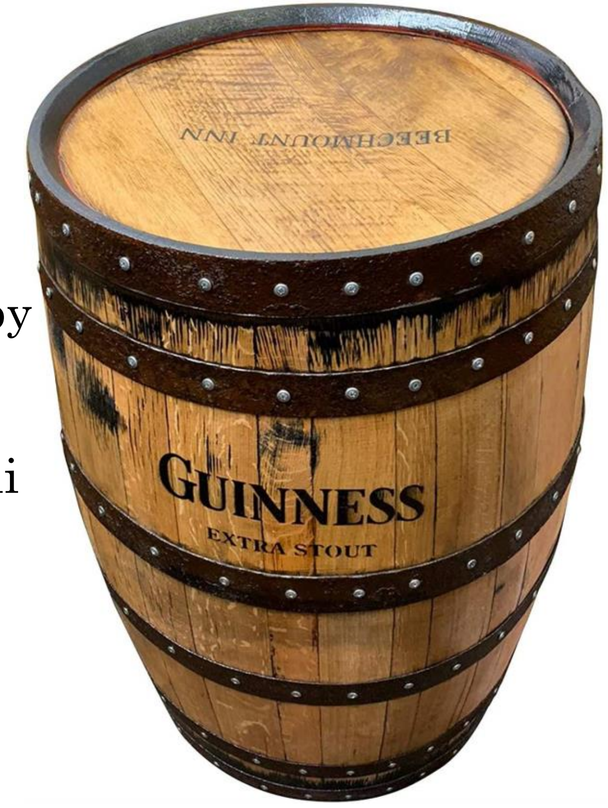
Natalia Józefacka

Wprowadzenie

- https://www.youtube.com/watch?v=bqfcFCjaE1c

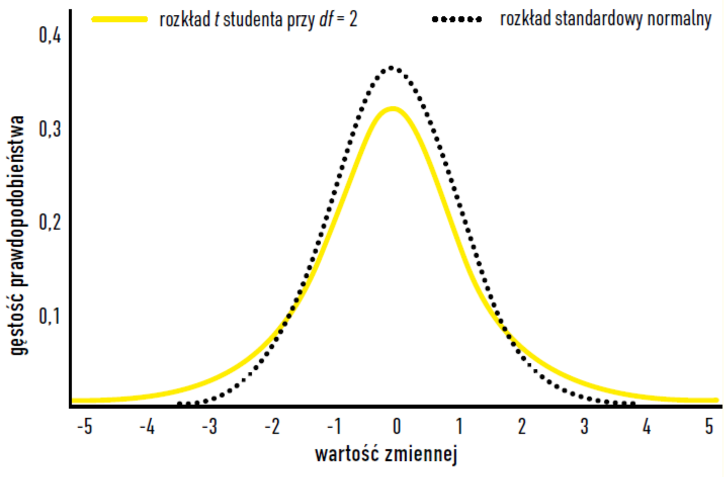
Test t jako rozwiązanie dla małych prób

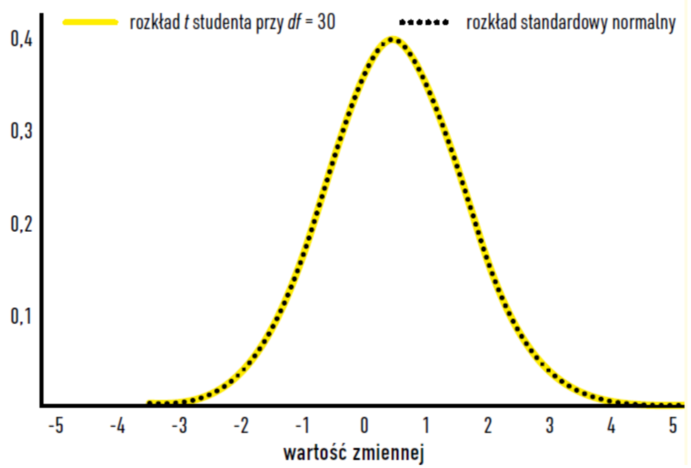
- Gosset – potrzebował testu dla $n<30$
- Chciał mieć pewność, że poziom ekstraktu słodowego (który stanowił dla kontroli jakości smaku kluczową rolę) mieści się w zakresie +/− 0,5 od docelowej wartości, która w Guinnessie wynosiła 133
- Na podstawie własnych obliczeń doszedł on do wniosku, że sprawdzenie losowo i wielokrotnie zaledwie 2-elementowej próby (czyli skosztowanie porcji z dwoch beczek piwa), a następnie sprawdzenie, ile znajduje się w nich słodu, dawało w 80% przypadków pomiar mieszczący się w wyznaczonej granicy (czyli $133 \pm 0{,}5$).



Rozkład *t* Studenta

- Jest to korekta do standardowego rozkładu normalnego
- Wraz ze wzrostem liczby stopni swobody rozkład t dąży do standardowego rozkładu normalnego. Zbieżność ta postępuje na tyle szybko, że już przy kilkudziesięciu stopniach swobody rozkład t jedynie minimalnie różni się od rozkładu statystyki Z





Rodzina testów t

1. Test *t* Studenta dla jednej próby → porównuje wynik jednej grupy obserwacji (np. uśrednionego poziomu słodu z kilku porcji piwa) (**I**) do znanego, założonego kryterium (poziom słodu w Guinnessie = 133) (**II**)

2. Test *t* Studenta dla prób niezależnych → porównuje wynik jednej, niezależnej grupy obserwacji (np. średniej ilości tygodniowego spożycia piwa w grupie w grupie studentów I roku) (**I**) do drugiej, niezależnej (czyli najzwyczajniej w świecie po prostu innej, rożnej, niezwiązanej z tą pierwszą) grupy obserwacji (np. średniej ilości tygodniowego spożycia piwa w grupie studentów V roku) (**II**)

3. Test *t* Studenta dla prób zależnych → porównuje wyniki z jednego pomiaru w jednej grupie (np. średnia ilość spożycia piwa wśród studentów na tydzień przed sesją) (**I**) do wyników z drugiego pomiaru w tej samej grupie (średnia ilość spożycia piwa wśród tych samych studentów tydzień po sesji) (**II**) lub wyniki, które można dobrać w pary i które, jak sama nazwa wskazuje, mogą być od siebie zależne (np. średnia ilość tygodniowego spożycia piwa przez ciebie – **I** i twojego/twoją współlokatora/kę – **II**, bo chyba wiesz, czemu to się ze sobą wiąże…).

Rodzina testów

To co ją łączy to:

- Porównywanie średnich
- Zawsze są 2 grupy/elementy

Mechanizm testu *t* Studenta

W celu policzenia testu potrzebne są:

- Wartości średnich (M) – obliczane na podstawie próby
- Wartość odchylenia (SD)
- Liczebność próby

Każdej wartości t (przy danej liczbie stopni swobody) odpowiada jakieś prawdopodobieństwo wystąpienia.

Dla takiego rozkładu wyznaczony jest jednocześnie **obszar krytyczny**, posiadający **z dwóch stron wartości krytyczne**, do ktorych porównuje się wyliczoną statystykę.

Ta z kolei, jeśli je przekroczy (tzn. będzie wyższa niż wartość krytyczna dodatnia lub niższa, jeśli wartość krytyczna jest ujemna),wskazuje na wynik istotny.

Podsumowując

- wyliczenie statystyki testowej *t* dla konkretnych wyników uzyskanych w analizie,
- sprawdzenie, czy znalazła się ona w obszarze krytycznym ustalonym dla danego rozkładu *t* Studenta o określonej liczbie stopni swobody i określonym kryterium istotności – najczęściej $\alpha = 0{,}05$ (inaczej mówiąc, czy przekroczyła ona wartość bezwzględną wartości krytycznej ustalonej dla tego rozkładu),
- sprawdzenie i podanie, jaka jest wyliczona dla niej **wartość prawdopodobieństwa *p***, które później porównuje się do założonego **kryterium istotności α**, podejmując tym samym decyzję dotyczącą hipotezy zerowej.

Hipotezy statystyczne test t dla 1 próby

- Przykład hipotezy badawczej, w ktorej badacz zakłada kierunek: *Średni poziom słodu w losowo wybranych porcjach piwa jest niższy od poziomu słodu we wzorcowym Guinessie.*

- Hipotezy statystyczne:

- H_0 (zerowa): *Średni poziom słodu w losowo wybranych porcjach piwa nie różni się od poziomu słodu we wzorcowym Guinessie.*

- H_1 (alternatywna bezkierunkowa): *Średni poziom słodu w losowo wybranych porcjach piwa różni się od poziomu słodu we wzorcowym Guinessie.*

Hipotezy statystyczne
test *t* dla prób niezależnych

- Przykład hipotezy badawczej, w której badacz zakłada kierunek: *Średnia tygodniowa ilość piwa spożywana w grupie studentów V roku jest mniejsza od średniej tygodniowej ilości piwa spożywanej w grupie studentów I roku.*

- Hipotezy statystyczne:

- H_0 (zerowa): *Średnia tygodniowa ilość piwa spożywana w grupie studentów V roku nie różni się od średniej tygodniowej ilości piwa spożywanej w grupie studentów I roku.*

- H_1 (alternatywna bezkierunkowa): *Średnia tygodniowa ilość piwa spożywana w grupie studentów V roku różni się od średniej tygodniowej ilości piwa spożywanej w grupie studentów I roku*.

Hipotezy statystyczne test *t* dla prób zależnych

Przykład hipotezy badawczej, w ktorej badacz zakłada kierunek: *Średnia ilość piwa spożywana w grupie studentów tydzień przed sesją jest mniejsza od średniej ilości piwa spożywanej tydzień po sesji*.

Hipotezy statystyczne:

H_0 (zerowa): *Średnia ilość piwa spożywana w grupie studentów tydzień przed sesją nie różni się od średniej ilości piwa spożywanej tydzień po sesji*.

H_1 (alternatywna bezkierunkowa): *Średnia ilość piwa spożywana w grupie studentów tydzień przed sesją różni się od średniej ilości piwa spożywanej tydzień po sesji*.

Założenia testu

Założenia

Dotyczy wszystkich trzech wariantów testu:

- Pomiar na skali ilościowej – warunkiem jest możliwość obliczenia średniej.
- Porównanie dwóch średnich.
- Rozkład normalny – dotyczy rozkładu próbkowania (UWAGA! Jeżeli ilość obserwacji przekracza 30 osob195, nie musisz bezwzględnie sprawdzać tego założenia!).

Dotyczy tylko wariantu dla prób niezależnych:

- **Homogeniczność** wariancji – zmienność w obydwa grupach powinna być z reguły na podobnym poziomie (UWAGA! Można robić odstępstwa, ale trzeba zadbać o inne kryteria, np. rozkład i rownoliczność; w przypadku złamania tego założenia można stosować poprawkę na nierowne wariancję).
- **Równoliczność grup** – liczebność nie musi być identyczna, ale zbliżona.

Homogeniczność wariancji

- Założenie dotyczy że wariancja w obu porównywanych grupach powinna być podobna
- Sprawdza się za pomocą testu Levena, jeżeli jest nieistotny, oznacza to że wariancje nie różnią się znacząco od siebie, a więc są homogeniczne
- Istotny test Levena wskazuje na heterogeniczność wariancji, wtedy trzeba odczytywać wynik testu z poprawką

Równoliczność grup

- Nie oznacza że grupy mają mieć dokładnie tyle samo obserwacji TYLKO, że ilość osób między grupami nie może się istotnie różnić
- Chi kwadrat dobroci dopasowania – nieistotny oznacza, że grupy są równoliczne
  - istotny że grupy różnią się od siebie – wtedy warto zastosować test nieparametryczny *U* Manna-Whitneya

Wnioskowanie

Gdybyś po użyciu **testu dla jednej próby** otrzymał wynikprawdopodobieństwa *p* = 0,437, to nie miałbyś podstaw do odrzucenia hipotezy zerowej (wynik jest większy od założonego przez nas kryterium α = 0,05, a zatem jest nieistotny).

Skoro brzmiała ona: *Średni poziom słodu w losowo wybranych porcjach piwa nie różni się od poziomu słodu we wzorcowym Guinessie*, a na podstawie wartości *p* nie można jej podważyć, to znaczy, że te porcje piwa, których smakowałeś, faktycznie nie rożnią się (a konkretnie nie rożni się średni poziom słodu) od wzorca, a zatem – HURRA – trafiłeś na dobry sort!

Co by było, gdyby jednak wynik *p* okazał się mniejszy niż α = 0,05 (np. *p* = 0,037)? Musiałbyś wówczas odrzucić hipotezę zerową, a w konsekwencji przyjąć alternatywną

(Średni poziom słodu w losowo wybranych porcjach piwa różni się od poziomu słodu we wzorcowym Guinessie).

W tym momencie byłbyś również zobligowany do opisania tych rożnic, tj. określenia, czy w tym przypadku średnia, jaką otrzymałeś, jest większa czy mniejsza od kryterium wynoszącego 133.

Wnioskowanie

Przy istotnym wyniku dla **testu *t* dla prób niezależnych** (np. $p = 0{,}001$) odrzuciłbyś hipotezę zerową *(Średnia tygodniowa ilość piwa spożywana w grupie studentów z V roku nie różni się od średniej tygodniowej ilości piwa spożywanej w grupie studentów I roku) i uznałbyś, że między grupami różnice istnieją, wskazując jednocześnie (na podstawie wielkości średniej)*, że to np. w grupie pierwszoroczniaków spożycie jest większe.

Gdyby wynik wynosił np. $p = 0{,}764$ ($p > 0{,}050$), musiałbyś pokornie stwierdzić, że nie masz podstaw do odrzucenia H_0 i w twoim badaniu różnic między grupami nie udało się wykazać.

Wnioskowanie

W teście dla prób zależnych uzyskanie np. wyniku $p = 0{,}049$ ($p < 0{,}050$) oznaczałoby, iż po odrzuceniu hipotezy zerowej mówiącej o braku różnic należałoby przyjąć alternatywą (Średnia ilość piwa spożywana w grupie studentów tydzień przed sesją *różni się od średniej ilości piwa spożywanej tydzień po sesji*), a następnie wskazać (na podstawie wielkości średniej), iż np. spożycie piwa w grupie studentów wzrasta po sesji.

W przypadku $p > 0{,}050$ (np. $p = 0{,}056$) hipotezy zerowej nie mógłbyś uznać za fałszywą i doszedłbyś do wniosku, że średnie spożycie piwa wśród studentów przed i po sesji wydaje się plasować na podobnym poziomie (innymi słowy badanie akurat tej próby nie dało podstaw do uznania, że różnice istnieją, choć wcale nie jest oczywiste, że w populacji ich faktycznie nie ma!).

Siła efektu

- *d* Cohena - Współczynnik ten jest obliczany na podstawie różnicy średnich między grupami (w przypadku prób niezależnych) lub pomiarami, czy parami wyników (w przypadku prób zależnych), która jest dzielona przez wielkość odchylenia standardowego.

Współczynnik *d* Cohena	Interpretacja
0,2–0,5	Słaby efekt
0,6–0,7	Umiarkowany efekt
0,8–1,3	Silny efekt
>1,4	Bardzo silny efekt

Siła efektu

Jeżeli d Cohena wynosi 0,5, oznacza to, że średnie dwóch porównywanych grup różnią się o 0,5 odchylenia standardowego. Oznacza to, że wartość przeciętnej osoby w grupie pierwszej wynosi o poł (0,5) odchylenia standardowego więcej niż w grupie drugiej.

Siła efektu

Odsetek osób w grupie 2 który byłby poniżej średniego wyniku osoby w grupie 1.

Wartość *d* Cohena	Procent osób w grupie 2, która miałaby wyniki poniżej przeciętnego wyniku osoby z grupy 1
0,0	50%
0,2	58%
0,4	66%
0,6	73%
0,8	79%
1,0	84%
1,2	88%
1,4	92%
1,6	95%
1,8	96%
2,0	98%
2,5	99%
3,0	99,9%

ANOVA

Wprowadzenie

- https://www.youtube.com/watch?v=HgcJ07heUI0 (w domu)
- https://www.youtube.com/watch?v=ZOxHtdDjAB0
- https://www.youtube.com/watch?v=TUVb4oV99cs

Wprowadzenie

- Sir Ronald Fisher
- Rozszerzenie testu t Studenta
- **ANOVA** (**AN**alysis **O**f **VA**riance).

Jednoczynnikowa ANOVA

- 1 czynnik = 1 zmienna która dzieli na grupy
- Posiada więcej niż 2 unikatowe wartości

- Hipotezy statystyczne
- H_0: *Średni poziom energii nie różni się w zależności od rodzaju spożywanego napoju.*
- H_1: *Średni poziom energii różni się w zależności od rodzaju spożywanego napoju.*

- Rodzaje napojów: woda, yerba mate, energetyk, kawa

Wariancja wewnątrzgrupowa

- **Wariancja wewnątrzgrupowa**: to średni poziom wariancji w analizowanych grupach

Grupa	*n*	*M*	*sd*	*Var*
Woda	60	25,87	3,27	10,71
Yerba Mate	60	23,07	3,31	10,94
„Energetyk"	60	17,05	2,78	7,70
Kawa	60	9,58	3,36	11,30

W tym wypadku wynosi 10,16

Wariancja międzygrupowa

- **Wariancja międzygrupowa**: w odróżnieniu od poprzedniej mówi o rozproszeniu wyników pomiędzy poszczególnymi grupami (czyli o ile odchylają się średnie z poszczególnych grup od średniej ogólnej dla wszystkich grup razem).
- Obliczenie wykonuje się w kilku krokach

1) Oblicz ogólną średnią (grand mean) czyli średnią ze średnich

Grupa	*n*	*M*	*sd*	*Var*
Woda	60	25,87	3,27	10,71
Yerba Mate	60	23,07	3,31	10,94
„Energetyk"	60	17,05	2,78	7,70
Kawa	60	9,58	3,36	11,30

W tym wypadku wynosi 18,89

Wariancja międzygrupowa

2) Od poszczególnych średnich odejmujesz *grand mean* i podnosisz do kwadratu aby usunąć ujemne wartości

3) Uzyskany wynik mnożysz przez liczebność grupy

krok2 krok3

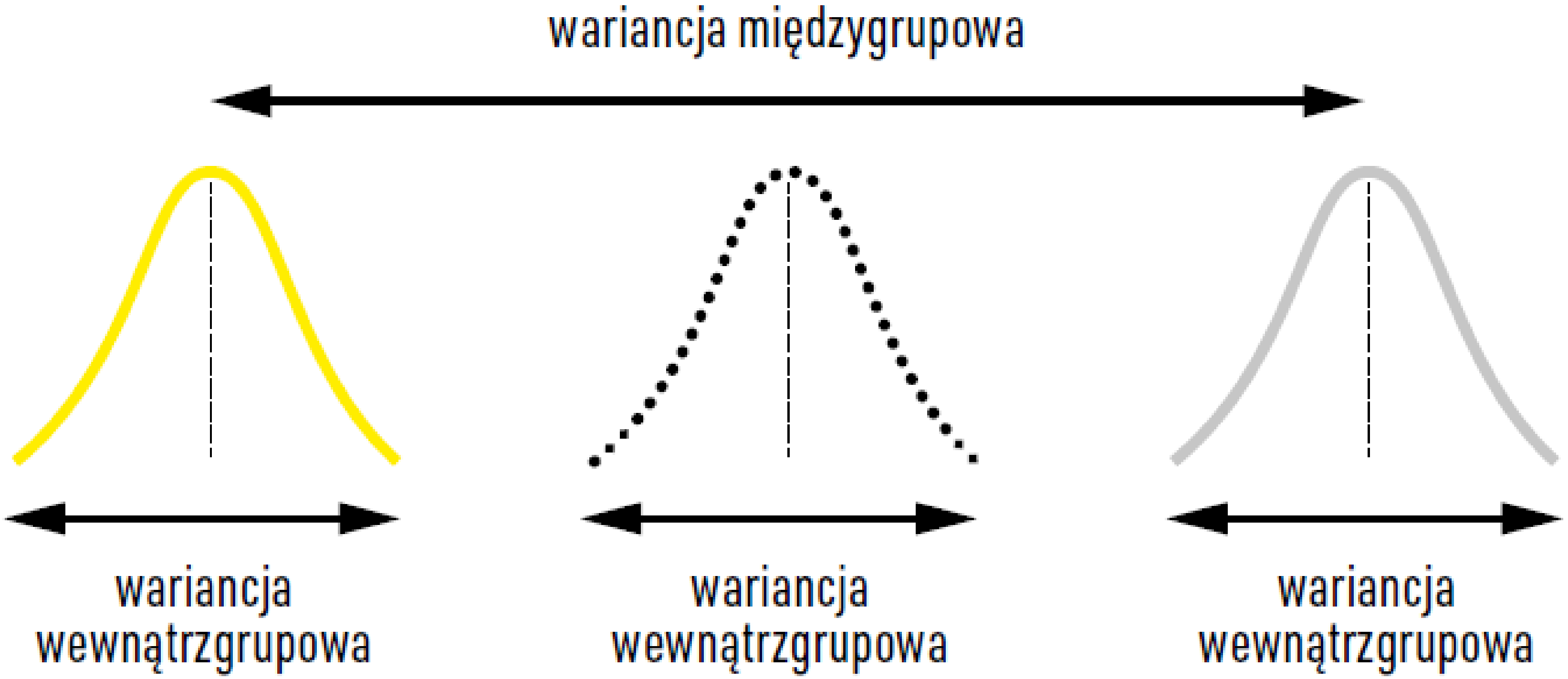
Grupa	M	n	$(M - M_{og})$	$(M - M_{og})^2$	$(M - M_{og})^2 * n$
Woda	25,87	60	6,98	48,72	2923,20
Yerba Mate	23,07	60	4,18	17,47	1048,20
„Energetyk"	17,05	60	-1,84	3,39	203,40
Kawa	9,58	60	-8,91	86,68	5200,80
Średnia	18,89			Suma	9375,60

Wariancja międzygrupowa

Uzyskaną w ten sposób sumę dzielisz przez liczbę grup minus 1 i otrzymujesz współczynnik wariancji międzygrupowej:

$$\frac{9375,60}{4-1} = 3125,20$$

Wariancja wew- i międzygrupowa



Mechanizm testu

- Wariancje są potrzebne do wyliczenia wartości statystyki F
- ANOVA porównuje wariancję wewnątrzgrupową z wariancją międzygrupową
- Zależy nam aby
  - Wariancja wewnątrzgrupowa – była jak najmniejsza
  - Wariancja międzygrupowa – była jak największa

Mechanizm testu

Ponieważ wariancja opisuje rozrzut wyników wokół średniej, na podstawie **wysokich wartości wariancji międzygrupowej** można powiedzieć, że średnie z grup mają duży rozrzut wokół średniej ogólnej. Ich wyniki są więc od siebie istotnie rożne.

Jednak aby różnice te mogły wybrzmieć, istotnym warunkiem jest również **niska wartość wariancji wewnątrzgrupowej** (wynikająca z faktu, iż grupy są bardzo spójne wewnętrznie i jednorodne – cechują się małą zmiennością wyników w swoim obrębie).

W przeciwnym razie wysoki rozrzut wyników wewnątrz grup uniemożliwia wyciągnięcie wiarygodnych wniosków co do średnich w poszczególnych grupach.

Stosunek wariancji międzygrupowej do wariancji wewnątrzgrupowej nazywa się statystyką *F*, na cześć sir Ronalda Fishera, twórcy analizy wariancji:

$$F = \frac{\text{wariancja}_{\text{międzygrupowa}}}{\text{wariancja}_{\text{wewnątrzgrupowa}}}$$

Przykład

- W omawianym wcześniej przykładzie z rodzajem spożywanego napoju i poziomem energii:

$$F = \frac{3135,20}{10,16} = 307,60$$

- Nie wiadomo jednak, czy te 307,60 to dużo czy mało.
- Aby to określić, należy znać **wartość krytyczną**, do której możesz porównać tę otrzymaną w badaniu.
- W celu uzyskania odpowiedzi, czy wartość 307,60 przekracza wartość krytyczną, analizuje się rozkład statystyki F o określonej **liczbie stopni swobody**.

Przykład

- Statystyka F ma dwa rodzaje stopni swobody

1. oblicza się poprzez odjęcie od liczebności całej próby liczby grup, co w tym przypadku daje: $df_1 = 4 - 1 = 3$.
2. uzyskuje się poprzez odjęcie liczby grup od całkowitej liczby obserwacji. Tak więc w przypadku omawianego badania: $df_2 = 240 - 4 = 236$ (wynika to z tego, że w każdej z 4 grup jest jeden wynik, który nie ma już żadnej swobody jeśli chodzi o swoją potencjalną wartość).

WAŻNE

- jeśli statystyka F jest mniejsza od 1, oznacza to, że wariancja niewyjaśniona (wewnątrzgrupowa) jest większa od wariancji wyjaśnionej (międzygrupowej).
- Odnosi się to do sytuacji, w której występują duże, losowe, niezaplanowane (i dlatego niewyjaśnione) różnice wewnątrz badanych grup, a stosunkowo niewielkie różnice pomiędzy nimi. Mówi się wówczas o braku efektu.
- Jeśli współczynnik F jest większy od 1, można dopiero domniemywać, że wynik będzie istotny statystycznie, niemniej zależy to oczywiście od tego czy różnice faktycznie istnieją, jaka jest liczebność obserwacji, a także jakie zostało ustalone kryterium istotności.
- Te dwie ostatnie własności powodują, że kształt rozkładu statystyki F zmienia się, a co za tym idzie zmieniają się również wartości krytyczne.

Założenia testu

Założenia

Dotyczy wszystkich trzech wariantów testu:

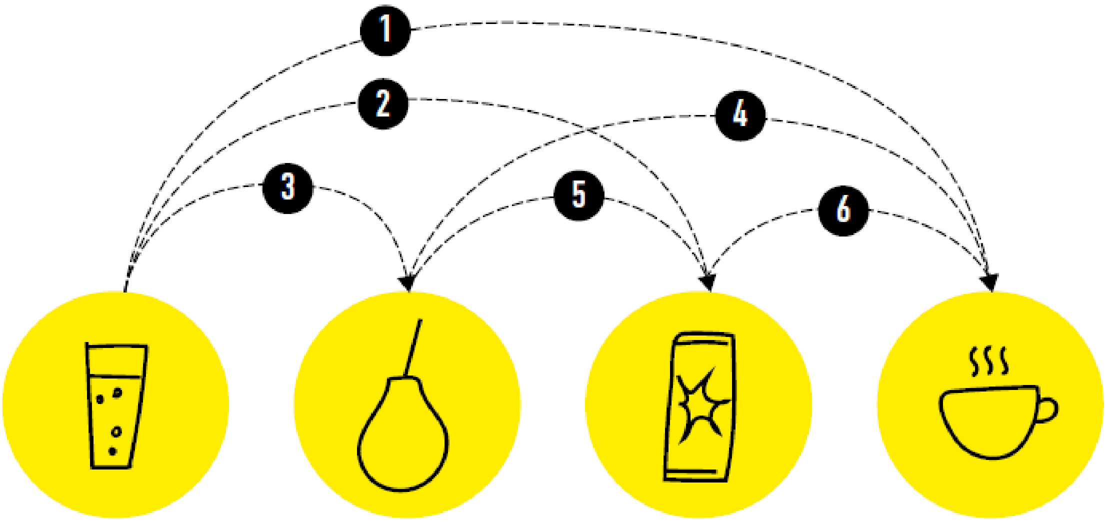
- Pomiar na skali ilościowej – warunkiem jest możliwość obliczenia średniej.
- Porównanie dwóch średnich.
- Rozkład normalny – dotyczy rozkładu próbkowania (UWAGA! Jeżeli ilość obserwacji przekracza 30 osob195, nie musisz bezwzględnie sprawdzać tego założenia!).

Dotyczy tylko wariantu dla prób niezależnych:

- **Homogeniczność** wariancji – zmienność w obydwa grupach powinna być z reguły na podobnym poziomie (UWAGA! Można robić odstępstwa, ale trzeba zadbać o inne kryteria, np. rozkład i rownoliczność; w przypadku złamania tego założenia można stosować poprawkę na nierowne wariancję).
- **Równoliczność grup** – liczebność nie musi być identyczna, ale zbliżona.

Testy post-hoc

- Istotny wynik testu F wskazuje, że są różnice, jednak żeby dowiedzieć się między którymi grupami należy zastosować testy post-hoc
- Porównują one parami średnie i wskazują które dokładnie się między sobą różnią



Równe wariancje	
Nazwa testu	**Opis**
NIR (LSD)[229]	Testy liberalne; brak kontroli błędu I rodzaju
SNK[230]	Testy liberalne; brak kontroli błędu I rodzaju
Bonferroni	Testy bardzo konserwatywne, ścisła kontrola błędu I rodzaju
Tukey	Testy bardzo konserwatywne, ścisła kontrola błędu I rodzaju
Sidak	Testy konserwatywne, ścisła kontrola błędu I rodzaju
Scheffe	Testy konserwatywne, ścisła kontrola błędu I rodzaju
REGWQ[231]	Testy umiarkowanie konserwatywne; najlepiej używać, kiedy masz równe wariancje i liczebności grup
Gabriel	Najlepiej używać, kiedy masz równe wariancje i niewiele różniące się liczebności grup
Hochbergs GT2	Najlepiej używać, kiedy masz równe wariancje i mocno różniące się liczebności grup

Nierówne wariancje	
Nazwa testu	**Opis**
Tamhanes T2	Odporne na nierówne wariancje, ale zbyt konserwatywne
Dunnets T3	Odporne na nierówne wariancje, ale zbyt konserwatywne
Dunnets' C	Odporne na nierówne wariancje, ale zbyt konserwatywne
Games-Howell	Idealny dla grup o nierównej wariancji

Siła efektu

- Do wyboru masz trzy232 rożne miary: **ε2 (epsilon), ω2 (omega) i η2 (eta)**.
- Za obciążoną najmniej uznaje się statystykę ε2.
- Parametr ω2 zazwyczaj jest niedoszacowany,
- η2 w większości przypadków jest przeszacowany i ukazuje zawyżoną siłę danego efektu w populacji.

Wartość miary	Siła efektu
0,00-0,01	Bardzo słaba
0,02-0,04	Słaba
0,05-0,16	Umiarkowana
0,17-0,36	Relatywnie silna
0,37-0,64	Silna
0,65-1,00	Bardzo silna

Analiza regresji liniowej jednozmoennowej

Wprowadzenie

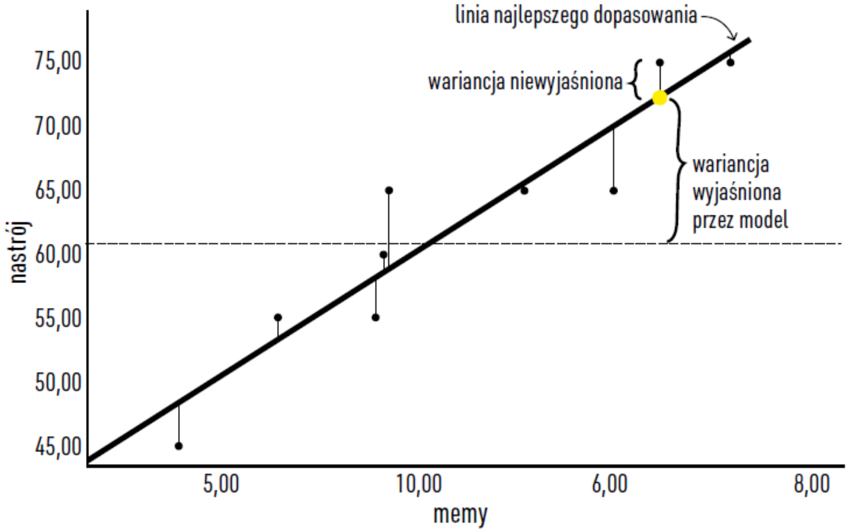
- Powstała na podstawie korelacji
- Podstawę stanowi równanie regresji czyli rodzaj funkcji liniowej
- Umożliwia predykcje zmiany w obrębie jednej zmiennej na podstawie drugiej zmiennej
- Wykorzystuje 2 rodzaje zmiennych:
  - Zmienna predykcyjna (wyjaśniająca, predyktor)
  - Zmienna wyjaśniana (kryterium)

Równanie regresji

- Równanie regresji – wzór funkcji liniowej
- *Y = a * X + b*
- Dzięki niemu właśnie badacz jest w stanie przewidzieć wartość Y na podstawie wartości X, a co istotne, wartości obydwu tych zmiennych mogą być zupełnie dowolne
- Liczby ustalone: a i b (tzw. parametry regresji)
    - wartości możesz poznać dopiero po zebraniu danych na temat zmiennej X i Y.
    - a – miejsce przecięcia z osią OY (inaczej wyraz wolny lub odcięta)
    - b – współczynnik kierunkowy regresji, informuje o nachyleniu linii regresji. Wykorzystywany jest do zebrania informacji o ile jednostek przeciętnie wzrośnie lub zmaleje wartość zmiennej wyjaśnianej Y, gdy wartość predyktora X wzrośnie o jedną jednostkę.

Metoda najmniejszych kwadratów i linia najlepszego dopasowania

- Metoda najmniejszych kwadratów pozwala na przeprowadzenie linii przez całą chmurę punktów na wykresie, a czyni to w taki sposób, aby odległość między każdym punktem a przebiegającą linią była możliwie jak najmniejsza. Uzyskaną w ten sposób prostą nazywa się **linią najlepszego dopasowania**



Linie pionowe wyrysowane od prostej do poszczególnych punktow to błędy oszacowania lub **reszty** regresji, które pokazują, jak bardzo rzeczywiste (empirycznie uzyskane) wyniki różnią się od wyrysowanej linii prostej (wyników wyliczonych na podstawie równania).

Celem metody najmniejszych kwadratów jest znalezienie takiego nachylenia linii i jej kierunku (czyli takich współczynników regresji), które sprawią, że suma podniesieniowych do kwadratu reszt jest jak najmniejsza.

Gauss, karłowata planeta Ceres i metoda najmniejszych kwadratów

- Carl Friedrich Gauss (1777-1855)
- obliczyć sumę 1 + 2 + 3.. aż do 100
  - skrót polegający na sparowaniu liczb (1+100) + (2 + 99) + …. (50+51) w ten sposób uzyskał 50 par, równych 101, wynik wyniósł równe 5050 – a rozwiązanie nazywamy współcześnie serią arytmetyczną.
- w 1801 r. odkryto karłowatą planetę Ceres
- Gauss rozważając ten problem założył:
  - Małe błędy są bardziej prawdopodobne niż duże.
  - Prawdopodobieństwo wielkości błędów jest równe, a ich rozkład symetryczny.
  - Gdy wykonuje się kilka pomiarów, najbardziej prawdopodobną wartością jest średnia.
- Krzywa dzwonowa = prawo częstości błędu
- Książe matematyki

https://www.actuaries.digital/2021/03/31/gauss-least-squares-and-the-missing-planet/

Hipotezy

- Hipotezy statystyczne:
- H0 (zerowa): *Nie można przewidzieć poziomu nastroju za pomocą zmiennej „liczba oglądanych memow”* (współczynnik regresji kierunkowy dla zmiennej „liczba oglądanych memow” rowna się 0).
- H_1 (alternatywna bezkierunkowa): *Na podstawie wartości zmiennej „liczba oglądanych memow” można przewidzieć poziom nastroju* (współczynnik regresji kierunkowy dla zmiennej „liczba oglądanych memow” jest rożny od 0).

ANOVA w regresji

- wyliczając za pomocą metody najmniejszych kwadratów współczynnik regresji kierunkowy (czyli *b*) dla zmiennej *X*, nie jesteś w stanie określić, czy jest on istotnie różny od 0 czy nie.
- Do tego służy ANOVA – test F
- w regresji o większej liczbie zmiennych predykcyjnych, istotność każdej z nich analizowana jest za pomocą testu *t*, a analiza wariancji (test *F*) służy globalnemu testowaniu istotności całego modelu.
- Jednakże w przypadku regresji prostej wynik testu *t* i analizy wariancji dają identyczne wartości. Co więcej, wielkość statystyki *F* jest w tej sytuacji równa drugiej potędze wartości statystyki *t* (a wartość prawdopodobieństwa *p* dla nich obydwu jest taka sama).

ANOVA w regresji

Ogólną funkcją analizy wariancji przeprowadzanej dla regresji prostej jest odpowiedź na pytania:

1) czy model jest dobrze dopasowany do danych (czyli czy teoretyczne wyniki przewidywań zmiennej *Y* na podstawie *X* uzyskane dzięki równaniu regresji pokrywają się z wynikami empirycznymi), a także,

2) czy model uwzględniający zmienną *X* jest lepszy w prognozowaniu wartości *Y* niż sama tylko średnia tej zmiennej.

- Żeby odpowiedzieć sobie na powyżej postawione dwa pytania, za pomocą analizy wariancji porównywana zostaje wariancja resztowa (niewyjaśniona przez model) do wariancji wyjaśnionej przez regresję.

$$F = \frac{\text{wariancja}_{\text{regresji}}/df_1}{\text{wariancja}_{\text{reszty}}/df_2}$$

ANOVA w regresji

- Uzyskana w ten sposób statystyka *F* jest porównywana z wartością krytyczną rozkładu *F* dla hipotezy zerowej, dzięki czemu zwracana zostaje wartość *p* (czyli poziom prawdopodobieństwa otrzymania takiej wartości *F* przy założeniu, że hipoteza zerowajest prawdziwa).
- Jeśli jest ona mniejsza niż przyjęty poziom istotności, pozwala to na stwierdzenie, że model jest dobrze dopasowany do danych, a zawarte w modelu zmienne predykcyjne istotnie lepiej przewidują zmienną *Y* niż sama średnia.
- wynik istotny wskazuje, iż wartości obserwacji uzyskane w badaniu leżą bliżej wyrysowanej linii najlepszego dopasowania, aniżeli średniego wyniku zmiennej wyjaśnianej.

ANOVA w regresji

- Co ważne, analiza wariancji podaje także inne miary niezbędne do oszacowania **współczynnika determinacji *R*2**.
- Miara ta pozwala na określenie, jaki procent zmienności wyników zmiennej Y (np. poziomu nastroju) tłumaczony jest wynikami zmiennej predykcyjnej *X* (np. liczbą obejrzanych memów).
- Jej wartość możliwa jest do ustalenia dzięki określeniu proporcji wariancji dla regresji (a zatem tej wyjaśnionej) do wariancji ogólnej (suma wariancji wyjaśnionej i wariancji niewyjaśnionej / resztowej), którą następnie mnoży się przez 100 (aby uzyskać wartość procentową).
- Innym sposobem obliczenia tego wskaźnika jest podniesienie do kwadratu miary korelacji między zmiennymi (*R*)

Współczynnik standaryzowany *Beta*

- $Y = a * X + b$
- Współczynnik b jest trudny do interpretacji
- Beta – jest wystandaryzowanym współczynnikiem b
- Oblicza się ją ustalając stałą a – 0 oraz wyrażając b w jednostkach odchylenia standardowego

Założenia analizy regresji

Założenia:

• Zmienne na **skali ilościowej**270

• Dane są **sparowane**

• Obserwacje **niezależne**

• Odpowiednia liczba obserwacji – *n* musi być większe lub równe liczbie szacowanych parametrów,

Diagnostyka rozkładu

Prostoliniowość między zmiennymi – dwie zmienne są ze sobą powiązane w taki sposób, że zmianie wartości jednej zmiennej towarzyszą przewidywalne zmiany drugiej zmiennej.

Wartości **odstające** – uważaj szczególnie na te wpływowe!

Diagnostyka reszt

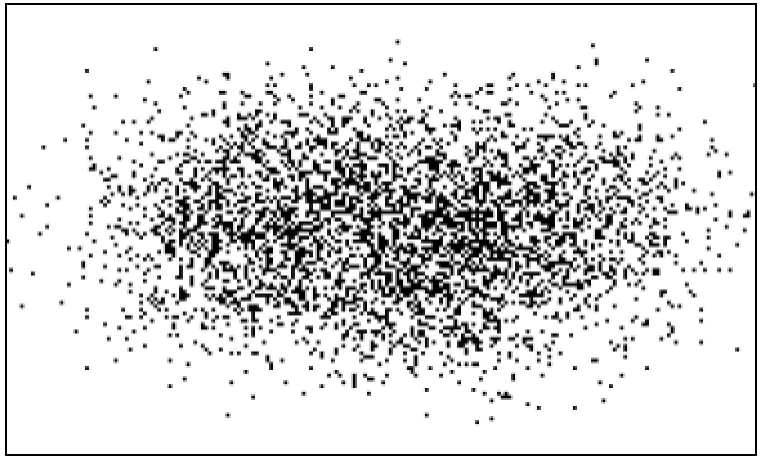
- **Homoscedastyczność** – trudno brzmiące słowo mowiące o rownym rozłożeniu wariancji reszt wokoł linii regresji.

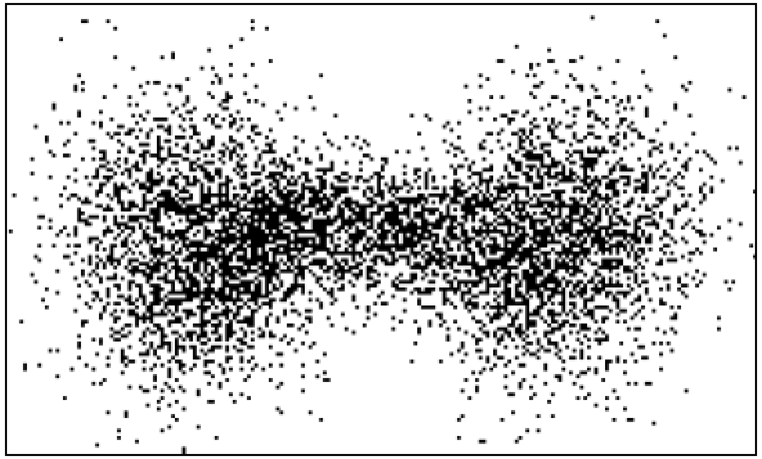
- Rozkład **normalny** – tym razem dotyczy **rozkładu reszt** (**UWAGA! Ważne przy małych liczebnościach proby**).

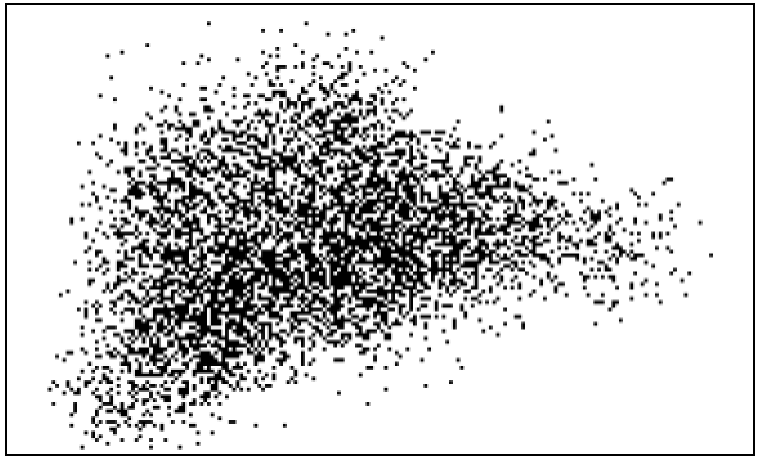
- **Brak autokorelacji** – **UWAGA! Założenie ważne przy analizie szeregow czasowych** (dot. pomiaru zmiennych zależnych).

Homoskedastyczność

- Rozkład wariancji wokół linii regresji powinien być równomierny
- Przy niedopilnowaniu tego założenia otrzymacie błędne oszacowanie parametrów modelu







Rozkład normalny reszt

- **Założenie o normalności dotyczy tu jednak nie rozkładu probkowania zmiennej wyjaśniającej czy wyjaśnianej, ale rozkładu reszt.**
- Złamanie założenia o normalności rozkładu reszt powoduje wzrost prawdopodobieństwa popełnienia błędu I rodzaju, a co za tym idzie szacowany przez program współczynnik istotności statystycznej może okazać się niedokładny
- prawidłowość ta obowiązuje jedynie w sytuacji, kiedy liczebność twojej próby jest bardzo mała (bezpiecznie jest przyjąć bezwzględną granicę poniżej 20 obserwacji).

Wnioskowanie

- Współczynnik determinacj R^2
- Istotność modelu – statystyka F
- Ważność predyktorów – beta

Dziękuję