

 **Od zmienności ciągłej do identyfikacji genów**

Kornelia Polok i Roman Zieliński
Biologia molekularna, kierunek lekarski
2023/2024

Od zmienności ciągłej do identyfikacji genów

- 1. Cechy ilościowe**
 - Definicja
 - Właściwości cech ilościowych
- 2. Układ doświadczalny**
 - Populacja i próba
 - Randomizacja i blokowanie
 - Błędy
- 3. Metody statystyczne**
 - Rozkład częstości
 - Miary tendencji
 - Miary rozprożeń
 - Testowanie hipotez
- 4. QTL**
 - Definicja QTL
 - Mapowanie QTL



„Statystyka wydobywa prawdę z chaosu”



1. Cechy ilościowe: definicja

Cechy ilościowe to cechy mierzalne (masa, ilość). Uwarunkowane są wieloma genami, z których każdy ma niewielki wpływ na fenotyp.



Selekcja sztuczna: selekcja w kierunku cech użytkowych, np. duże owoce.



Selekcja naturalna: selekcja osobników najlepiej przystosowanych do środowiska. Zwiększa ogólne przystosowanie populacji.

Identyfikacja genów warunkujących cechy ilościowe wymaga zastosowania zaawansowanych metod statystycznych i molekularnych.

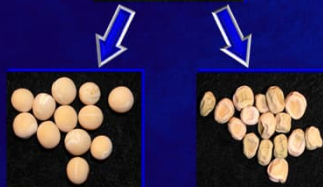
Cechy ilościowe stanowią większość obserwowanej zmienności. Cechy te najczęściej podlegają selekcji w toku ewolucji, a także w procesach hodowlanych.



1. Cechy ilościowe: definicja

Analiza cech ilościowych jest utrudniona ze względu na brak jednoznacznych kategorii – klas fenotypowych.

Cechy jakościowe



Każdego osobnika można jednoznacznie zaklasyfikować do danej kategorii, np. nasiona gładkie i pomarszczone.

Cechy ilościowe

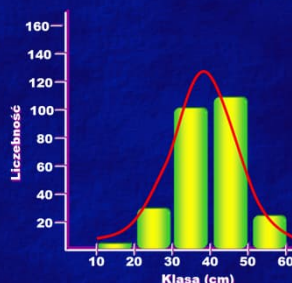


Występuje całe spektrum fenotypów, trudno je zaklasyfikować do klas, np. rośliny są wysokie, nieco niższe, średnie, niskie, bardzo niskie itd. Klasy wyodrębnia się arbitralnie, np. co 10 cm (klasa 11-20, 21-30, 31-40 itd).



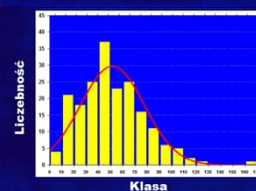
1. Cechy ilościowe: właściwości

Cechy ilościowe są kodowane przez wiele genów co skutkuje występowaniem wielu fenotypów.



Wysokość roślin jako przykład cechy ilościowej. U człowieka do cech ilościowych zalicza się: wzrost, masę ciała, barwę skóry, kolor włosów.

Poligeny: wiele genów zlokalizowanych na różnych chromosomach ale współdziałających ze sobą w wytwarzaniu cechy fenotypowej. Odpowiadają one QTL (Quantitative Trait Loci).



Masa zielona, F₂ HU x BO

Efekt każdego genu jest mały i dlatego różnice pomiędzy fenotypami są niewielkie, co obserwuje się jako zmienność ciągłą, rozkład cech jest zbliżony do rozkładu normalnego (krzywa Gaussa).

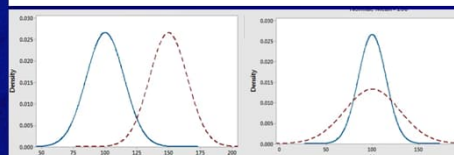
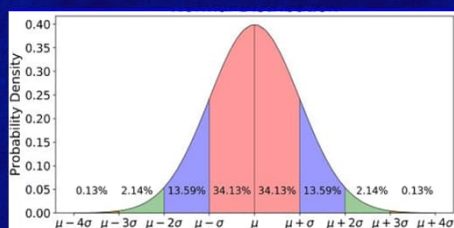


1. Cechy ilościowe: właściwości

Zmienność fenotypowa cechy ilościowej w populacji ma charakter ciągły, zbliżony do rozkładu normalnego.

Rozkład normalny

- Ma wiele cech ułatwiających wprowadzenie testów statystycznych.
- Łatwo skalowany, gdy cecha nie ma typowego rozkładu normalnego.
- Odchylenie standardowe determinuje wysokość i szerokość wykresu.
- Obszar pod wykresem zawsze wynosi 1.
- Może dawać wartości ujemne dla średniej bliskiej zeru, co jest nierealistyczne dla większości danych biologicznych.



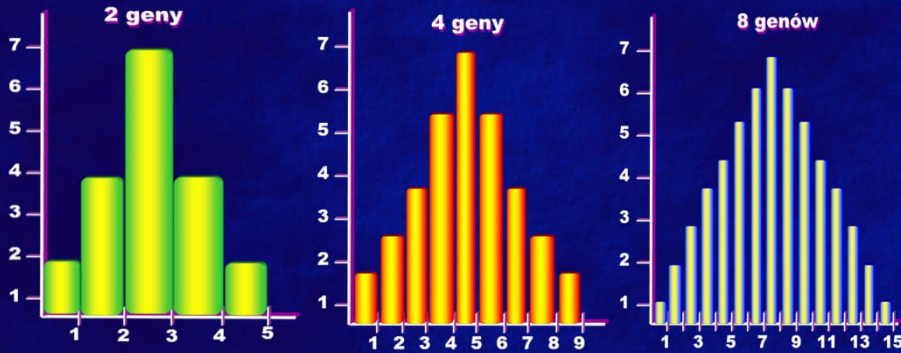
Rozkład normalny oraz wpływ wartości średniej (lewa) oraz odchylenia standardowego (prawa) na kształt wykresu

Rozkład normalny (Gaussa): funkcja średniej i wariancji, maksimum osiąga dla wartości średniej, następnie opada symetrycznie w obu kierunkach.



1. Cechy ilościowe: właściwości

Im więcej genów odpowiada za daną cechę tym więcej powstaje klas fenotypowych.



Przykładowa liczba klas fenotypowych w zależności od liczby genów.

Przy dużej liczbie genów wpływających na cechę, efekty wywołane przez pojedyncze geny mogą być mniejsze od zmian powodowanych przez środowisko.



1. Cechy ilościowe: właściwości

Wzrost człowieka jest typową cechą, która może mieć zarówno charakter jakościowy jak i ilościowy.

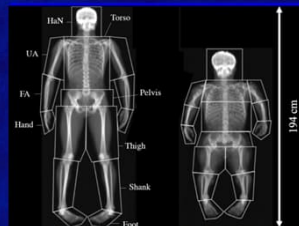
Czynniki wpływające na zmienność wzrostu człowieka jako cechy ilościowej

■ Czynniki genetyczne:

- wiele segregujących genów,
- niewielki wpływ danego genu na genotyp;
- współdziałanie genów.

■ Czynniki środowiskowe:

- zależność ekspresji genów od środowiska;
- brak jednoznacznego związku między genotypem a fenotypem;
- różnice w zależności od wieku, diety itd.



Jeżeli gen ma znaczny efekt fenotypowy to obserwuje się typowe dziedziczenie mendelowskie. Tak dziedziczy się achondroplasia, która jest uwarunkowana mutacją w genie *FGFR*. Mutacja jest dominująca.

Jeżeli efekt fenotypowy genu jest nieznaczny i wiele genów warunkuje cechę wówczas obserwuje się rozkład normalny, np. wzrost człowieka.



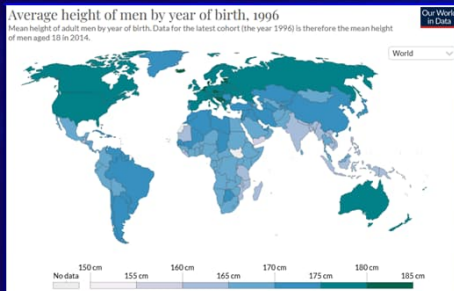
Statystyka w analizie cech ilościowych ma na celu dostosowanie stosunkowo prostych funkcji matematycznych do danych.



1. Cechy ilościowe: właściwości

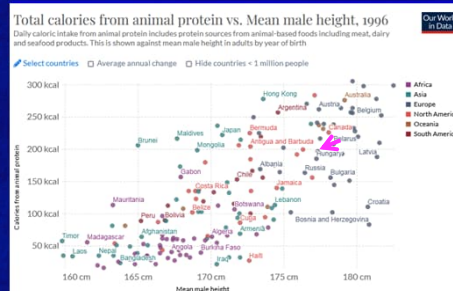
Wzrost człowieka wykazuje silną zmienność środowiskową – koreluje z szerokością geograficzną i rodzajem diety.

Zróżnicowanie geograficzne wzrostu populacji ludzkich.



Średni wzrost mężczyzn na świecie wynosi 171 cm, kobiet 159 cm (rocznik 1996). Wzrost wykazuje zróżnicowanie geograficzne, niższe osoby są w Południowej Azji, wyższe w Europie i Azji Centralnej.

Wpływ diety na wzrost populacji ludzkich na świecie



Istnieje pozytywna korelacja między jakością dostarczanego białka (na ogół w postaci produktów mięsnych) a wzrostem.

Dziedziczenie ilościowe dotyczy odporności na choroby, zaburzeń psychicznych oraz odpowiedzi na farmakoterapię.

Our World in Data 2022

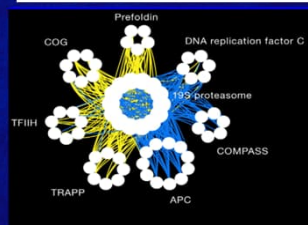
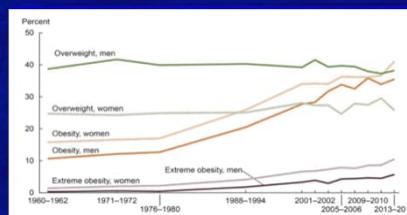


1. Cechy ilościowe: właściwości

Mechanizmy genetyczne leżące u podłoża cech ilościowych nie różnią się od tych warunkujących cechy jakościowe.

Genetyka cech ilościowych wyjaśnia:

- w jakim stopniu obserwowana zmienność wynika z różnic w genotypie, a w jakim jest uwarunkowana czynnikami środowiskowymi;
- ile genów odpowiada za cechę ilościową;
- jaki jest wkład pojedynczego genu;
- jak geny współdziałają ze sobą;
- jak szybko zmienia się cecha pod wpływem selekcji;
- jaka metoda selekcji jest najlepsza.



Nadwaga osób dorosłych w USA (góra). Sieć interakcji genetycznych w komórkach drożdży (dół).

Różnice pomiędzy cechami jakościowymi i ilościowymi dotyczą metod analizy, które uwzględniają wiele genów wpływających na daną cechę.



Od zmienności ciągłej do identyfikacji genów

1. Cechy ilościowe

- Definicja
- Właściwości cech ilościowych

2. Układ doświadczalny

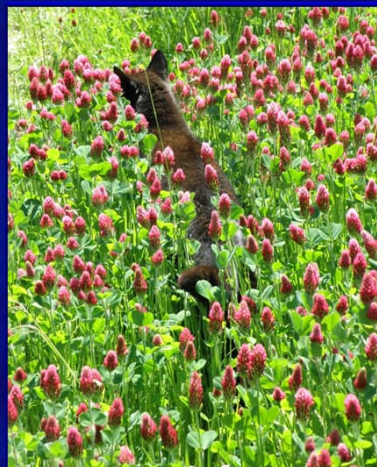
- Populacja i próba
- Randomizacja i blokowanie
- Błędy

3. Metody statystyczne

- Rozkład częstości
- Miary tendencji
- Miary rozprożeń
- Testowanie hipotez

4. QTL

- Definicja QTL
- Mapowanie QTL



„Statystyka wydobywa prawdę z chaosu”

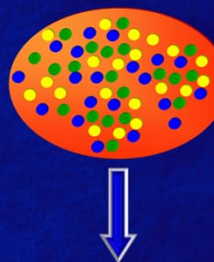


2. Układ doświadczalny: populacja i próba

Populacja statystyczna: zbiór elementów danego typu, np. grupa osobników o określonych cechach.

Populacje i próby

- Populacja skończona: liczba elementów zbioru jest skończona, najczęściej występuje w eksperymentach medycznych.
- Populacja nieskończona: liczba elementów zbioru jest bardzo duża, nieskończona, na ogół występuje w badaniach populacji naturalnych.
- Badania całej populacji są badaniami wyczerpującymi.
- Badanie części populacji to badanie cząstkowe przeprowadzone na podstawie próby statystycznej.



Badana populacja: grupa osobników danego gatunku, odmiany, ekotypu itd. Za populację można uznać np. zestaw genotypów, linii, pokolenie F_2 .

Próba: grupa osobników na podstawie której wnioskujemy o populacji.

Próba musi być reprezentatywna, dobór elementów musi być losowy, próba musi być odpowiednio liczna.



2. Układ doświadczalny: populacja i próba

Minimalna wielkość próby zależy od wielu czynników, w tym od wpływu danej cechy na fenotyp.

Wielkość próby

- Zbyt mała próba uniemożliwia sprawdzenie zgodności z rozkładem normalnym.
- Rozkład małych prób często odbiega od rozkładu normalnego.
- Małe próby uniemożliwiają wyciąganie jednoznacznych wniosków.
- Zaleca się tworzenie powtórzeń oraz odpowiedniej liczby kontroli. Im więcej powtórzeń tym mniejszy błąd doświadczenia. Liczba kontroli może być zwiększona, gdy z przyczyn niezależnych (np. rzadka choroba) populacja jest mała.

Różnica [%]	Szacowana wielkość próby
5	1450-3200
10	480-820
30	80-100
40	50-60

Wielkość próby w zależności od wielkości różnic fenotypowych (na podstawie Slutsky 2013).

Mała próba pozwala na wykrycie tylko dużych efektów. W małych próbach zjawiska losowe mogą maskować działanie badanego czynnika.



2. Układ doświadczalny: randomizacja i...

Blokowanie, powtórzenia i randomizacja minimalizują wpływ czynników losowych, pozwalają ocenić błąd doświadczenia.

Układ bloków randomizowanych

- Umożliwia odróżnienie efektów wynikających z działania różnych czynników, np. traktowania i genotypu.
- Możliwe jest obliczenie interakcji pomiędzy działającymi czynnikami, np. efekt leku zależy od genotypu.
- Umożliwia podział zmienności całkowitej na środowiskową i genetyczną oraz podział zmienności genetycznej na poszczególne komponenty.
- Pozwala wnioskować o mechanizmach działania genów.



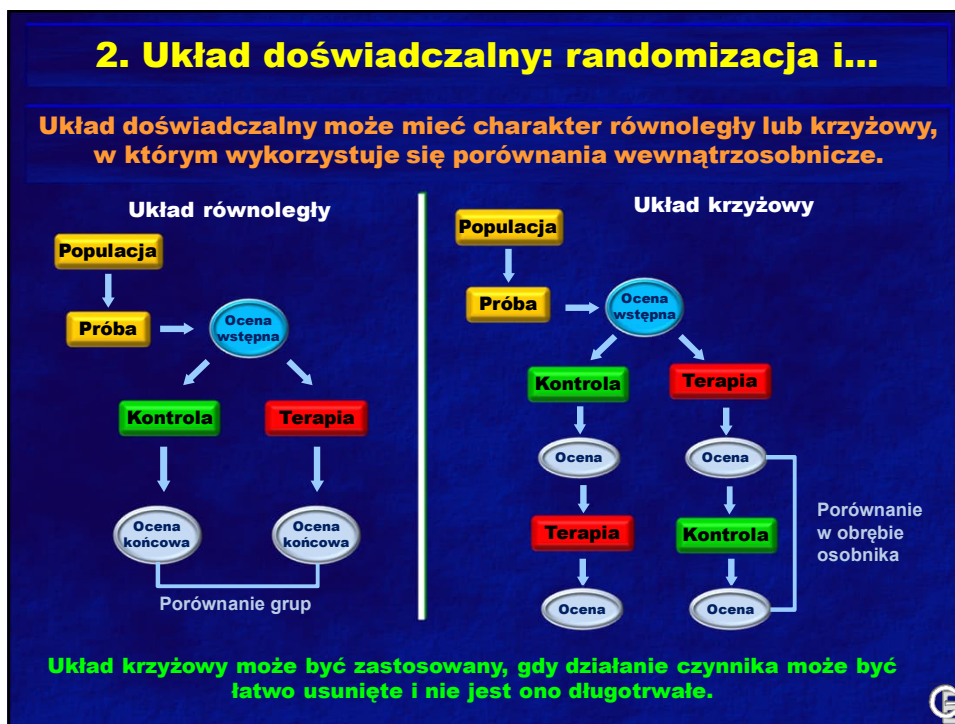
Układ bloków randomizowanych: składa się z bloków i powtórzeń.

Powtórzenie: np. ten sam genotyp powtórzony w różnych blokach. Na jedno powtórzenie składa się wiele (np. 30) osobników.

Bloki: w każdym bloku występują wszystkie badane genotypy, kolejność jest losowa, w każdym bloku inna. W bloku może występować jedno lub więcej powtórzeń.

RCT: randomizowane doświadczenia kliniczne, pojęcie dotyczy tylko przydziału osób do grup, a nie układu bloków i powtórzeń.





2. Układ doświadczalny: randomizacja i...

Badania kohortowe (populacyjne): obserwacja uczestników w czasie w celu oceny wpływu jakiegoś czynnika środowiskowego itp.

Badania kohortowe: zasady i problemy

- Mogą być stałe (stała liczba uczestników) lub dynamiczne (uczestnicy mogą się zmieniać).
- Grupa musi być reprezentatywna dla populacji. Warunek trudny do uzyskania, gdyż często wybiera się np. określoną grupę zawodową.
- Jeżeli analiza obejmuje czynnik ryzyka to zmienia się on w czasie.
- Jeżeli wiele osób rezygnuje, to wartość badania spada.



Badania kohortowe są drogie ze względu na czas trwania, trudno zapewnić powtarzalne warunki.



2. Układ doświadczalny: błędy

Próba kontrolna jest często zbyt mała lub nieprawidłowo wprowadzona.

Nieprawidłowości związane z kontrolą

- Grupa kontrolna nie jest poddawana interwencji, np. przez podanie placebo.
- Grupa kontrolna różni się od grupy badanej, np. wpływ treningu jest analizowany u sportowców w stosunku do ogółu społeczeństwa.
- Grupa kontrolna jest świadoma spodziewanych wyników (np. brak efektów podania leków, gdy wiadomo, że lek wywołuje np. nudności), co może wpływać na jej zachowanie.



Nieprawidłowy układ: kontrola nie jest poddawana „zerowej” interwencji



Prawidłowy układ: kontrola jest poddana interwencji „zerowej”

Wszystkie czynniki, poza analizowanym, powinny działać identycznie na próbę i grupę kontrolną. Grupa kontrolna musi być poddawana podobnym procedurom jak grupa badana.



2. Układ doświadczalny: błędy

Częstym błędem jest tworzenie układu jednoczynnikowego zamiast wieloczynnikowego.

Błędny układ doświadczenia

- **Układ I:** w analizie działania leku na populację myszy wykorzystano dwa szczepy. Układ doświadczalny obejmował cztery grupy: K1, P1, K2 i P2. Cała zmienność jest objęta zmiennością między grupami. Nie wiadomo jaki wpływ ma szczep i interakcja.
- **Układ II:** źródłem zmienności jest: szczep, traktowanie, interakcja szczep x traktowanie. Możliwa analiza zależności efektu leku od szczepu (interakcja).



Układ jednoczynnikowy uniemożliwia wyciąganie wniosków o działaniu poszczególnych czynników i interakcji.



Od zmienności ciągłej do identyfikacji genów

1. Cechy ilościowe

- Definicja
- Właściwości cech ilościowych

2. Układ doświadczalny

- Populacja i próba
- Randomizacja i blokowanie
- Błędy

3. Metody statystyczne

- Rozkład częstości
- Miary tendencji
- Miary rozproszeń
- Testowanie hipotez

4. QTL

- Definicja QTL
- Mapowanie QTL



„Statystyka wydobywa prawdę z chaosu”



3. Metody statystyczne

Statystyka odgrywa kluczową rolę w planowaniu doświadczeń oraz w podejmowaniu decyzji zdrowotnych.

Statystyka w artykułach naukowych*

- 91,7% artykułów zawierało statystykę opisową i podstawowe tabele z danymi.
- 61,6% artykułów prezentowało dane w postaci wykresów.
- 72% prac wykorzystywało wartość P i przedziały ufności.
- 18% prac korzystało z większej liczby poziomów istotności (0,1, 0,05, 0,01).
- Tylko 5,1% prac wykorzystało kilka testów.

(*Na podstawie analizy 2016 artykułów z 2014 r.)

Metoda	Procent [%]
ANOVA	10
Regresja liniowa	19
Regresja logistyczna	36
Regresja Poissona	7
GLM	6
PCA/Analiza czynnikowa	4

Wykorzystanie zaawansowanych metod statystycznych w pracach z zakresu nauk medycznych i nauk o zdrowiu. Wykres przedstawia najczęściej wykorzystywane metody.

Wydawnictwo Naukowe PWN

Statystyka jest podstawą medycyny opartej na faktach. Wykorzystuje się statystykę opisową i wnioskowanie statystyczne.

Hayat et al. 2016

3. Metody statystyczne

W publikacjach informacje o stosowanych metodach statystycznych często są niepełne lub błędne.

Przykłady błędów w opisie metod statystycznych

- 2011: w 50% artykułów z 5 najlepszych czasopism neurologicznych wykorzystano błędny test w celu porównania dwóch czynników.
- 2018: 46% artykułów dotyczących interwencji prenatalnych nie uwzględniło niezależnych obserwacji i wykorzystało błędne testy.
- W pracach brakuje opisu układu doświadczalnego oraz wykorzystanych metod statystycznych.

Informacja	Procent [%]
Statystyka F	5
Stopnie swobody	2
Wartość p	5
Wielkość próby	75

Procent artykułów medycznych podających dane niezbędne do kontroli prawidłowości przeprowadzenia analizy wariancji (ANOVA).

Wydawnictwo Naukowe PWN

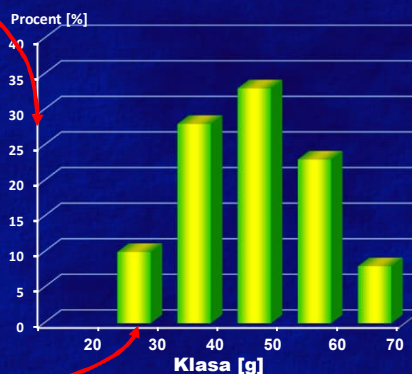
W pracach brakuje informacji jaki typ ANOVA jest wykorzystany, jaki typ testu post-hoc wykorzystano, brak informacji umożliwiających weryfikację.

Weissgerber et al. 2018.

3. Metody statystyczne: rozkład częstości

Rozkład częstości: opis populacji w postaci częstości osobników o określonym fenotypie.

Klasa [g]	Liczebność	Częstość
20-29	20	0,1 (10%)
30-39	55	0,28 (28%)
40-49	65	0,33 (33%)
50-59	45	0,23 (23%)
60-69	15	0,08 (8%)

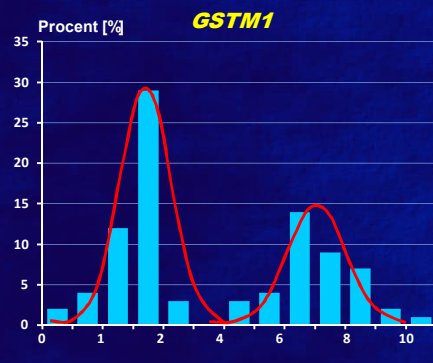


Histogram to graficzna prezentacja rozkładu. Rozkład normalny powstaje, gdy wiele niezależnych czynników wpływa na pomiar, geny działają addytywnie (efekty się sumują).

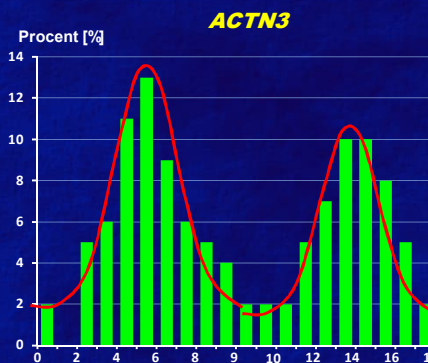


3. Metody statystyczne: rozkład częstości

Rozkład bimodalny powstaje, gdy na pomiar wpływają dwa czynniki lub dwie grupy czynników (np. dwie grupy genów, QTL).



Poziom ekspresji RNA dla genu *GSTM1* kodującego S-transferazę glutationową u 225 mężczyzn.



Poziom ekspresji RNA dla genu *ACTN3*, kodującego aktywninę alfa 3 u 225 mężczyzn.

Rozkład bimodalny dla poziomu ekspresji genów związanych z aktywnością mięśni wynikał z obecności e-QTL i polimorfizmu typu *cis*.

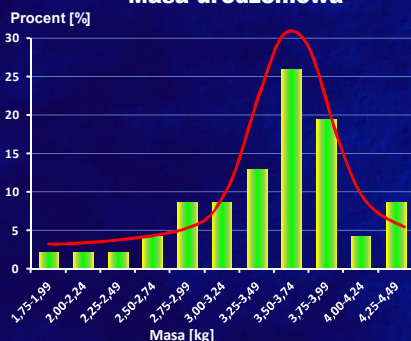
Mason et al., 2011



3. Metody statystyczne: rozkład częstości

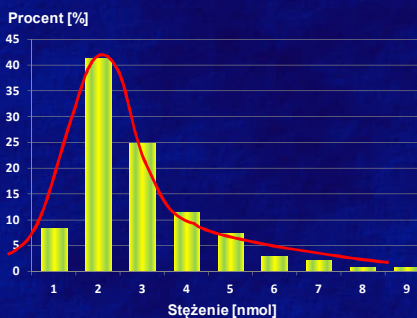
Rozkład asymetryczny powstaje, gdy dana cecha jest uwarunkowana współdziałaniem wielu czynników (np. wielu genów - epistaza).

Masa urodzeniowa



Asymetryczny wykres rozkładu masy urodzeniowej. Asymetria w kierunku wartości niższych.

Poziom trójglicerydów u mężczyzn



Asymetryczny wykres poziomu trójglicerydów u mężczyzn. Asymetria w kierunku wartości wyższych.

Rozkład asymetryczny można przekształcić w +/- normalny za pomocą skali logarytmicznej. Nie zawsze jest to polecane, gdyż na podstawie asymetrii można wnioskować o działaniu genów.

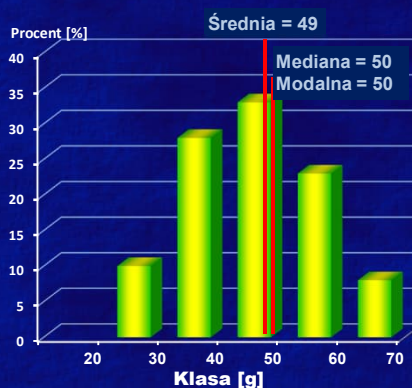


3. Metody statystyczne: miary tendencji

Miary tendencji centralnej służą do opisu populacji, charakteryzują badaną grupę, ale skupiają się na różnych aspektach.

Miary tendencji centralnej

- Średnia arytmetyczna: suma wszystkich pomiarów podzielona przez ich liczbę, algebraicznie zdefiniowana, znany rozkład.
- Mediana: wybiera pomiar środkowy (nieparzysta liczba) lub średnia z dwóch pomiarów środkowych (parzysta liczba), nie jest zdefiniowana algebraicznie, złożony rozkład.
- Modalna (dominanta): wartość, która powtarza się najczęściej, zwłaszcza użyteczna dla danych jakościowych, algebraicznie niezdefiniowana.



W przypadku rozkładu symetrycznego średnia, mediana i modalna dają podobne wartości.

Miary tendencji centralnej wyznaczają wielkość najbardziej charakterystyczną dla próby.

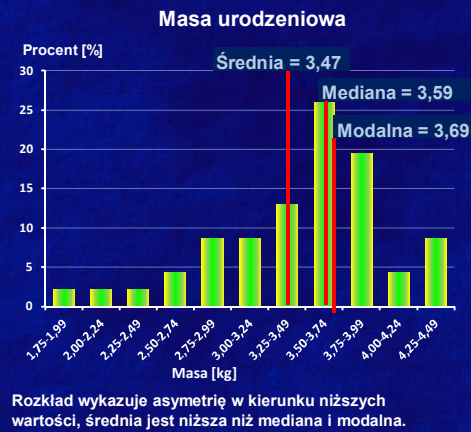


3. Metody statystyczne: miary tendencji

W rozkładach asymetrycznych mediana i modalna są przesunięte w stosunku do średniej w kierunku wartości niższych lub wyższych.

Znaczenie średniej arytmetycznej

- Średnia ma znaczenie w genetyce cech ilościowych, gdyż parametry genetyczne oraz rozproszenie alleli definiuje się w oparciu o średnią populacji.
- Obecność wyników znacznie odbiegających od większości sprawia, że średnia nie jest informatywna. W tej sytuacji lepiej sprawdza się mediana.
- Modalna jest użyteczna dla danych jakościowych lub zmiennych kategoryalnych (np. tak/nie).



Średnia jest szczególnie użyteczna w przypadku rozkładu normalnego. Wykorzystywana jest także w doświadczeniach wielopowtórzeniowych.



3. Metody statystyczne: miary tendencji

Średnia geometryczna ma zastosowanie gdy rozkład asymetryczny można przekształcić w rozkład logarytmicznie-normalny.

Średnią geometryczną n dodatnich liczb a_1, a_2, \dots, a_n nazywamy liczbę

$$G = (a_1 \cdot a_2 \cdot \dots \cdot a_n)^{1/n}$$

Przykład 1. Tempo wzrostu bakterii

- Szczep wzrasta o 20% po 1 h, 30% po 2 h i 50% po 3 h. Jeżeli początkowo było 100 bakterii to liczba bakterii:
 - po 1 h: 120, tempo wzrostu - 1,2;
 - po 2 h: 156, tempo wzrostu - 1,3;
 - po 3 h: 234, tempo wzrostu - 1,5.
- $G = (1,2 \cdot 1,3 \cdot 1,5)^{1/3} = 1,3276$, liczba bakterii po 3 h:
 - $100 \cdot (1,3276)^3 = 233,99$
- Średnia arytmetyczna: 1,33, liczba bakterii po 3 h - 235,26

Przykład 2. Tempo wzrostu drzewa

- Drzewo rośnie o 20% w 1 roku, 4% w 2 roku i 8% w 3 roku. Jeżeli początkowo drzewo miało 100 cm, to wysokość w latach wynosi:
 - 1 rok: 160, tempo wzrostu - 1,6;
 - 2 rok: 166, tempo wzrostu - 1,04;
 - 3 rok: 180, tempo wzrostu - 1,08.
- $G = (1,6 \cdot 1,04 \cdot 1,08)^{1/3} = 1,2158$, wysokość po 3 latach:
 - $100 \cdot (1,2158)^3 = 180$
- Średnia arytmetyczna: 1,24, wysokość po 3 latach - 191

Średnią geometryczną często stosuje się do oceny średniego tempa zmiany, np. wzrostu populacji.



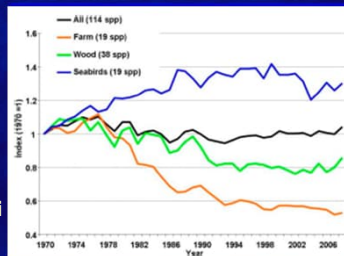
3. Metody statystyczne: miary tendencji

Średnia geometryczna, G jest wykorzystywana w ocenie względnej liczebności osobników w obrębie gatunków.

Wartość G w ocenie bioróżnorodności

- Średnia geometryczna lepiej oddaje trendy w zmianie względnej liczebności gatunków ponieważ zmiany liczebności mają charakter multiplikatywny (np. dwukrotny wzrost, spadek).
- Jeżeli liczebność wszystkich gatunków spada w jednakowym tempie, średnia geometryczna oddaje tę zmianę.
- Uzyskuje się takie same wyniki dla liczebności i biomasy.

Gatunek	1	2	3	X_{sr}	G
Rok 1	800	160	40	333	172
Rok 2	900	80	20	333	113
Zmiana	1,125	0,5	0,5	0,75	0,66



Zmiana względnej liczebności ptaków w ekosystemach na podstawie średniej geometrycznej w latach 1970-2006. Widoczny jest silny spadek bioróżnorodności w ekosystemach leśnych i agrokosystemach.

Względna liczebność osobników w gatunku jest wartością dynamiczną, która zmienia się w latach.

Buckland et al., 2011

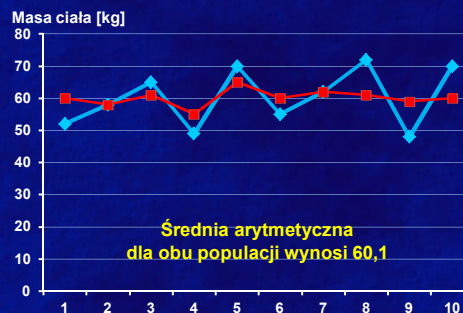


3. Metody statystyczne: miary rozproszenia

Miary rozproszenia opisują rozrzut pomiarów wokół miary tendencji centralnej.

Po co stosujemy miary rozproszenia?

- Różne zestawy danych mogą mieć taką samą średnią arytmetyczną, medianę, modalną.
- Im wyższa wartość miar rozproszenia tym bardziej wyniki są rozrzucone wokół tendencji centralnej. Przy dużym rozrzucie i zróżnicowaniu danych miary tendencji centralnych słabiej reprezentują wyniki.
- Przy niskiej wartości miar rozproszenia, miary tendencji centralnej są dobrymi reprezentantami próby.



Masa ciała w dwóch populacjach ludzkich. Średnia arytmetyczna jest taka sama (60,1 kg). Jednakże w populacji zaznaczonej na niebiesko rozrzut wyników jest znacznie większy.

Do najczęściej stosowanych miar rozproszenia zalicza się rozstęp (ang. range), odchylenie standardowe i wariancję.

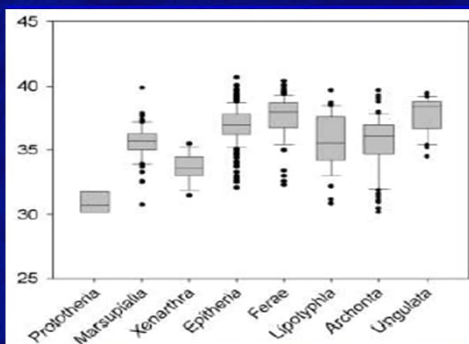


3. Metody statystyczne: miary rozproszenia

Rozstęp jest to różnica między wartością maksymalną i minimalną w próbie. Jest to najprostsza miara rozrzutu.

Rozstęp

- Dla wartości 7, 10, 8, 9, 5 rozstęp wynosi $10 - 5 = 5$.
- Czasami rozstęp przedstawia się jako zakres, 5-10.
- Im większa wartość rozstępu tym większy rozrzut danych.
- Rozstęp łatwo się oblicza, ale wykorzystywane są tylko dwie obserwacje i ma tendencję do wzrostu wraz ze zwiększeniem liczebności próby.
- Może być wykorzystany w badaniach filogenetycznych, w których porównuje się wiele grup.



Zakres temperatury ciała w °C u różnych grup ssaków i ptaków. Szare prostokąty wskazują zakres temperatur u grup, które zawierały co najmniej cztery gatunki.

Rozstęp jest miarą zawodną, gdyż różne rozkłady mogą mieć taki sam rozstęp.

Clarke | Rothery, 2007

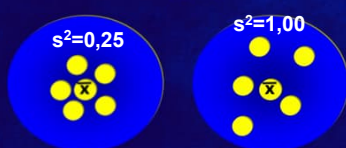


3. Metody statystyczne: miary rozproszenia

W praktyce do oceny rozproszenia najczęściej wykorzystuje się wariancję i odchylenie standardowe.

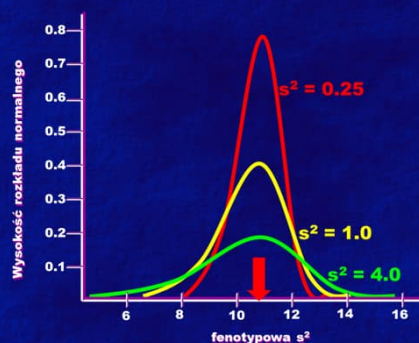
Wariancja jest to miara rozrzutu wartości wokół średniej. Jest definiowana jako odchylenie od średniego kwadratu średniej (MS)

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$



Te same średnie, ale większy rozrzut (wariancja) w populacji po prawej.

Odchylenie standardowe jest pierwiastkiem kwadratowym z wariancji. Jest podawane w tych samych jednostkach co oryginalne pomiary.



Dla tej samej średniej wygląd wykresu jest różny w zależności od wariancji.



3. Metody statystyczne: miary rozproszenia

W obliczeniu wariancji nie wykorzystuje się wielkości próby, ale liczbę tzw. stopni swobody (df).

No	Pomiar	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	112	8,7	75,69
2	102	-1,3	1,69
3	106	2,7	7,29
4	120	16,7	278,89
5	98	-5,3	28,09
6	106	2,7	7,29
7	80	-23,3	542,89
8	105	1,7	2,89
9	106	2,7	7,29
10	98	-5,3	28,09
\bar{x}	103,3	Suma	980,1
		Wariancja, s^2	108,9
		Odchylenie standardowe, s	10,44

Ponieważ obliczenie według sum kwadratów jest kłopotliwe w praktyce często oblicza się sumy pośrednie i wykorzystuje zależność:

$$S_T = S_a + S_e$$

> S_T : wariancja całkowita

> S_a : wariancja dla obiektów

> S_e : wariancja dla błęd

Współczynnik zmienności to stosunek odchylenia standardowego do średniej. Podaje się go w procentach: $10,44/103,3 * 100\% = 10,1\%$

Jeżeli dane pochodzą z populacji o rozkładzie normalnym, to $s = 10,44$ oznacza, że 68,3% populacji ma wartość cechy pomiędzy 93 ($103,3 - 10,44$) i 114 ($103,3 + 10,44$).

Liczba stopni swobody oznacza liczbę pomiarów niezależnych we wzorze. Ponieważ we wzorze na wariancję jest średnia, to $df = n - 1$.



3. Metody statystyczne: miary rozproszenia

Błąd standardowy średniej arytmetycznej określa dokładność obliczenia średniej populacji, dysponując średnią próby.

Błąd standardowy średniej arytmetycznej (SE)

- Dla każdej próby średnie arytmetyczne nieco się różnią.
- Średnie prób mają rozkład normalny z wariancją średniej i odchyleniem standardowym.
- Odchylenie standardowe dla średniej to błąd standardowy.
- Błąd standardowy określa rozproszenie serii średnich obliczonych w wielu próbach.
- $SE_x = s/n^{1/2}$ (odchylenie standardowe dzielone przez pierwiastek kwadratowy z liczby powtórzeń/pomiarów).

No	Pomiar	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	112	8,7	75,69
2	102	-1,3	1,69
3	106	2,7	7,29
4	120	16,7	278,89
5	98	-5,3	28,09
6	106	2,7	7,29
7	80	-23,3	542,89
8	105	1,7	2,89
9	106	2,7	7,29
10	98	-5,3	28,09
\bar{x}	103,3	Suma	980,1
		Wariancja, s^2	108,9
		Odchylenie standardowe, s	10,44
		Błąd standardowy, SE	$10,44/10^{1/2} = 10,44/3,16 = 3,3$

Błąd standardowy i odchylenie standardowe to różne pojęcia, które często są mylone. Błąd standardowy nie jest miarą rozproszenia.

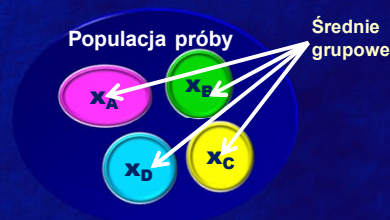


3. Metody statystyczne: testowanie hipotez

Analiza wariancji umożliwia porównanie wielu grup jednocześnie, umożliwia ocenę zmienności wewnątrz- i międzygrupowej.

Analiza wariancji (ANOVA)

- Wykorzystuje test F (Fishera-Snedecora) do oceny istotności różnic.
- Stosuje się w przypadku rozkładu normalnego lub rozkładu, który można przekształcić do normalnego.
- Wyróżnia się układ jednoczynnikowy, dwuczynnikowy, wieloczynnikowy.
- Istotna wartość testu F nie świadczy o występowaniu różnic między wszystkimi grupami.
- Stosuje się wraz z testami wielokrotnymi.



Analiza wariancji wykorzystuje spostrzeżenie, że w przypadku istotnego oddziaływania badanego czynnika na analizowaną cechę zmienność między grupami jest większa niż zmienność wewnątrz grup (tzw. błąd).

Całkowita zmienność badanej cechy może być wyrażona jako suma zmienności międzygrupowej i wewnątrzgrupowej.



3. Metody statystyczne: testowanie hipotez

Aby zweryfikować hipotezę zerową w analizie wariancji należy prawidłowo odczytać dane z tabeli analizy wariancji.

Źródło zmienności	ν liczba stopni swobody	SS Suma kwadratów	MS Średni kwadrat	Statystyka F	Wartość krytyczna*
Czynnik A	$a - 1$	SS_A	$SS_A/a - 1$	MS_A/MS_e	F dla ν_A, ν_e
Czynnik B	$b - 1$	SS_B	$SS_B/b - 1$	MS_B/MS_e	F dla ν_B, ν_e
A x B Interakcja	$(a - 1)(b - 1)$	SS_{AB}	$SS_{AB}/(a - 1)(b - 1)$	MS_{AB}/MS_e	F dla ν_{AB}, ν_e
Błąd (e)	$n - ab$	SS_e	$SS_e/n - ab$		
Całość	$n-1$				

- n: liczba wszystkich obserwacji
- a: liczba grup dla czynnika A
- b: liczba grup dla czynnika B

*Programy komputerowe często podają prawdopodobieństwo, wartość istotna jest zaznaczona na czerwono.



3. Metody statystyczne: testowanie hipotez

Przykład: wpływ stosowanego leku i fizjoterapii na ogólną ocenę aktywności choroby z reumatoidalnym zapaleniem stawów.

Czynnik 1	Czynnik 2	Samopoczucie
Fizjoterapia	Placebo	60
Fizjoterapia	Placebo	54
Fizjoterapia	Placebo	88
Fizjoterapia	Placebo	76
Fizjoterapia	Placebo	72
Fizjoterapia	Lek 1	48
Fizjoterapia	Lek 1	73
Fizjoterapia	Lek 1	39
Fizjoterapia	Lek 1	35
Fizjoterapia	Lek 1	51
Fizjoterapia	Lek 2	43
Fizjoterapia	Lek 2	67
Fizjoterapia	Lek 2	53
Fizjoterapia	Lek 2	48
Fizjoterapia	Lek 2	49
Brak F	Placebo	90
Brak F	Placebo	87
Brak F	Placebo	67
Brak F	Placebo	55
Brak F	Placebo	82
Brak F	Lek 1	56

Źródło zmienności	ν	SS	MS	F	Wartość krytyczna
Lek	3 - 1 = 2	1921,7	960,8	6,1	3,40
Fizjoterapia	2 - 1 = 1	974,7	974,7	6,1	4,26
Lek x Fizjoterapia	2	236,6	118,3	0,7	3,40
Błąd	30 - 3*2 = 24	3810,4	158,8		
Całość	30				

- W przykładzie uwzględniono tylko 5 pomiarów w grupie, jednakże w praktyce liczba pomiarów powinna być znacznie większa, np. 30 na grupę. Ponadto zastosowanie powtórzeń zwiększy dokładność oceny.
- Testowano dwa leki oraz kontrolę (placebo) – 3 grupy dla czynnika 1, leku.
- Testowano zastosowanie fizjoterapii lub jej brak – 2 grupy dla czynnika 2, fizjoterapii.
- Czerwoną kursywą zaznaczono wartości istotne statystycznie, tzn. obliczona wartość F jest wyższa od wartości granicznej.

Czynnikami, które istotnie wpływały na samopoczucie był lek oraz prowadzona fizjoterapia. Nie stwierdzono interakcji.



3. Metody statystyczne: testowanie hipotez

Istotna wartość testu F nie świadczy o istnieniu różnic pomiędzy wszystkimi obiektami.

Test LSD (N)					Test LSD (NIR)					
Test NIR, zmienna Plon (l-factor) Prawdopodobieństwa dla testów post-hoc Błąd: MS międzygrupowe = 43.397, df = 33.000					Test NIR, zmienna Plon (l-factor) Grupy jednorodnie, alfa = .05000 Błąd: MS międzygrupowe = 43.397, df = 33.000					
Nr podkl.	genotyp	{1}	{2}	{3}	Nr podkl.	genotyp	Plon Srednia	1	2	3
1	Mutant 1	66.167	56.500	71.000	4	Mutant4	55.1	****		
2	Mutant2	0.0104	0.0104	0.0003	2	Mutant2	56.5	****		
3	Mutant3	0.2127	0.0003	0.679	5	Kontrola	63.6		****	
4	Mutant4	0.0039	0.6791	0.0001	1	Mutant 1	66.2		****	****
5	Kontrola	0.4559	0.0297	0.0369	3	Mutant3	71.0			****

Test Tukeya						
HSD (nierówne N), zmienna Plon (l-factor) Przybliżone prawdopodobieństwa dla testów post hoc Błąd: MS międzygrupowe = 43.397, df = 33.000						
Nr podkl.	genotyp	{1}	{2}	{3}	{4}	{5}
1	Mutant 1	66.167	56.500	71.000	55.125	63.600
2	Mutant2	0.105794	0.105794	0.710471	0.047981	0.960531
3	Mutant3	0.710471	0.004971	0.004971	0.993389	0.221837
4	Mutant4	0.047981	0.993389	0.001904	0.001904	0.314447
5	Kontrola	0.960531	0.221837	0.314447	0.099117	0.099117

Przykładowe przedstawienie testów posthoc dla porównania różnych genotypów, wydruk z programu STATISTICA.

Pomiędzy wynikami dla różnych testów mogą wystąpić niewielkie różnice.

W celu oceny różnic pomiędzy poszczególnymi obiektami stosuje się testy wielokrotne (analiza posthoc)



Od zmienności ciągłej do identyfikacji genów

1. Cechy ilościowe

- Definicja
- Właściwości cech ilościowych

2. Układ doświadczalny

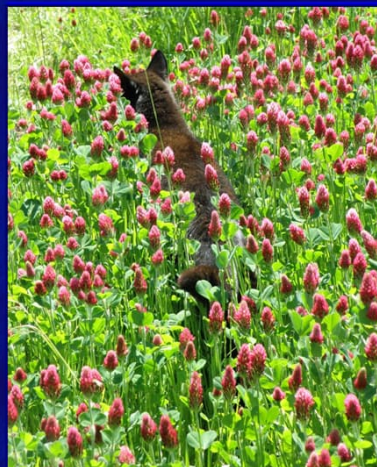
- Populacja i próba
- Randomizacja i blokowanie
- Błędy

3. Metody statystyczne

- Rozkład częstości
- Miary tendencji
- Miary rozprożeń
- Testowanie hipotez

4. QTL

- Definicja QTL
- Mapowanie QTL



„Statystyka wydobywa prawdę z chaosu”



4. QTL: geny warunkujące cechy ilościowe

QTL (ang. quantitative trait loci): gen lub grupa genów warunkująca cechę ilościową, zajmuje znaczny obszar chromosomu.

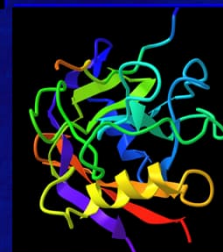
Poszukiwanie genów odpowiedzialnych za uzależnienie alkoholowe

- Populację segregującą otrzymano przez skrzyżowanie szczurów preferujących alkohol ze szczurami stroniącymi od alkoholu.
- Geny potencjalnie skorelowane z uzależnieniem wybrano na podstawie danych farmakologicznych oraz z doświadczeń z GMO.
- Analizowano ekspresję wybranych genów na poziomie mRNA.
- Analizowano sprzężenie z markerami o znanym położeniu.



Gen *Npy*, u człowieka chr. 7p15.3, koduje neuropeptyd Y, związany z uzależnieniami, u szczurów preferujących alkohol – niższy poziom ekspresji.

Gen *Sncα*, u człowieka chr. 4q22.1, koduje synukleinę alfa, białko obecne w neuronach, związane z chorobami neurodegeneracyjnymi, uczestniczy w regulacji syntezy dopaminy. U szczurów preferujących alkohol wyższa ekspresja.



O istotnej korelacji pomiędzy markerem a QTL wnioskuje się na podstawie testów statystycznych, np. testu t.

Spence et al., 2009

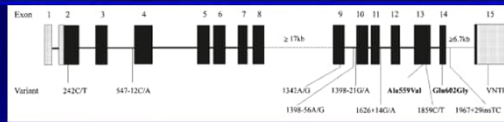


4. QTL: identyfikacja QTL

Mapowanie pojedynczych QTL umożliwiło powiązanie funkcji specyficznych regionów chromosomów z zaburzeniami ADHD.



Szczur SHR: nadaktywność ruchowa, impulsywność, deficyt uwagi, zmieniona neurotransmisja dopaminergiczna, zidentyfikowany polimorfizm w locus *DAT1*, który koduje transporter dopaminowy koreluje z objawami psychomotorycznymi. Badania z wykorzystaniem linii ssaków laboratoryjnych (szczury, myszy) umożliwiają identyfikację genów kandydatów odpowiedzialnych za cechy ilościowe człowieka.



Struktura *DAT1* człowieka na chr. 5p15.3. Na czarno zaznaczono egzony, szary fragment to tandemowe powtórzenia VNTR. Zmienność w obrębie VNTR koreluje z objawami neuropsychiatrycznymi.

Inne geny uczestniczące w powstaniu ADHD

- *SNAP25*: koduje białko z rodziny SNARE uczestniczące w transporcie synaptycznym, delecja w genie *SNAP25* u myszy powoduje nadpobudliwość ruchową.
- *TRβ* koduje receptor beta tarczycy, mutacja prowadzi do zespołu ADHD.

ADHD: zespół nadpobudliwości psychoruchowej z niedoborem uwagi, rozpoznawany u 3-6% dzieci w wieku szkolnym.



4. QTL: identyfikacja QTL

Mapowanie interwałowe polega na określeniu miejsca najbardziej prawdopodobnego wystąpienia QTL na mapie genetycznej.

W mapowaniu interwałowym:

- oblicza się średnie dla wszystkich klas fenotypowych;
- sprawdza się czy istnieją różnice między wartością cechy ilościowej dla wszystkich genotypów w locus markerowym;
- istotność różnic określa się na podstawie analizy wariancji;
- analizy prowadzone są dla każdej pary markerów przylegającej do QTL.



Obecność QTL jest przedstawiona graficznie jako funkcja LOD, czyli funkcja, która określa prawdopodobieństwo wystąpienia QTL w danym regionie chromosomu.

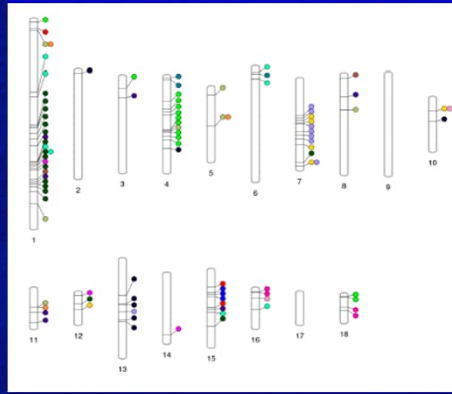
Interwał – region genomu zdefiniowany przez parę markerów, w którym prawdopodobieństwo położenia QTL jest największe.



4. QTL: identyfikacja QTL

Wysycione mapy genetyczne z dużą liczbą markerów w bliskiej odległości (~1 cM) są niezbędnym warunkiem mapowania QTL.

- Najwięcej QTL zmapowano u gatunków uprawnych, zwierząt domowych oraz u gat. modelowych.
- QTL mapuje się u człowieka.
- Mapowanie porównawcze u człowieka i świni domowej wykorzystywane jest do identyfikacji QTL odpowiedzialnych za otyłość:
 - zidentyfikowano 299 QTL dla 35 analizowanych cech;
 - QTL dla 20 cech znajdowały się w regionach kolinearnych z mapą genetyczną człowieka,
 - 19 regionów na mapie zawierało QTL dla kilku cech.



Lokalizacja QTL odpowiedzialnych za otyłość na mapie genetycznej świni domowej.

Pojedynczy QTL może zajmować duży obszar w chromosomie, nawet kilkadziesiąt centyMorganów (cM).

Pant et al., 2015

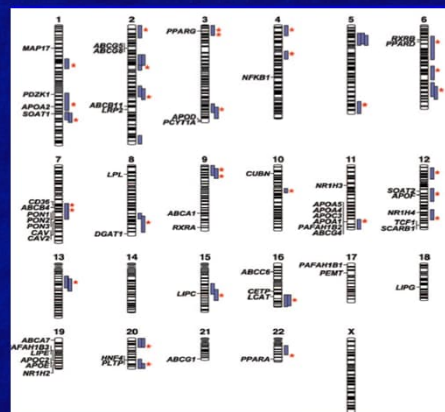


4. QTL: identyfikacja QTL

Około 30 QTL odpowiada za metabolizm cholesterolu HDL-C u człowieka. Regiony te są homologiczne do QTL u myszy.

Cholesterol HDL

- Cholesterol HDL zmniejsza ryzyko arteriosklerozy przez hamowanie modyfikacji cholesterolu LDL, redukcję chronicznego stanu zapalnego.
- Poziom cholesterolu HDL jest uwarunkowany genetycznie, odziedziczalność wynosi 40-60%.
- Geny uczestniczące w metabolizmie cholesterolu HDL kodują apolipoproteiny, białka transportujące, enzymy komórkowe, receptory i czynniki transkrypcyjne.



Mapa QTL dla cholesterolu HDL u człowieka.

Wyznaczone QTL wykorzystywane są do identyfikacji genów warunkujących daną cechę ilościową.

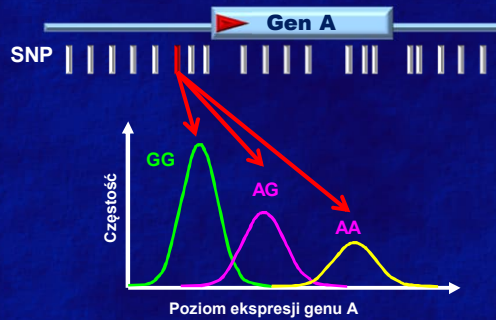


4. QTL: identyfikacja QTL

eQTL: QTL związany z różnicami w poziomie ekspresji danego genu, poszukuje korelacji między markerem a ekspresją.

Właściwości eQTL

- *cis* eQTL: związany z regionami w pobliżu genu (w odległości 1 Mbp od genu).
- *trans* eQTL: związany z regionami oddalonymi od genu, np. na różnych chromosomach.
- Różnice w eQTL występują w populacjach o różnym pochodzeniu, np. 17-29% genów w populacjach ludzkich ulega różnej ekspresji.
- Różnice w eQTL występują w zależności od tkanki, najczęściej dotyczą *cis* eQTL.



eQTL, poziom ekspresji genu zależy od genotypu SNP w pobliżu genu (czerwony). Najwyższa ekspresja genu A jest dla AA w SNP, najniższa - dla GG. Różnice w poziomie ekspresji genu A w zależności od genotypu SNP świadczą o regulacyjnym charakterze regionu zawierającego SNP.

Mapowanie eQTL umożliwia powiązanie polimorfizmu DNA i poziomu ekspresji genu z cechami ilościowymi.

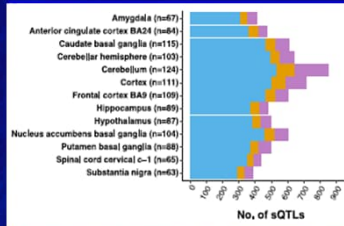


4. QTL: identyfikacja QTL

sQTL (splicing QTL): QTL regulujący alternatywną obróbkę (splicing) pre-mRNA, wykrywany za pomocą sekwencjonowania RNA.

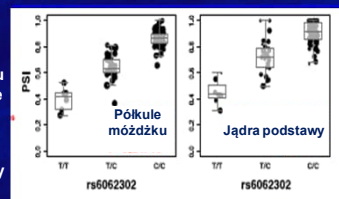
Zmienność sQTL w ludzkim mózgu

- 133 sQTL zidentyfikowano we wszystkich analizowanych (13) regionach ludzkiego mózgu.
- Od 387 do 849 sQTL było specyficznych dla poszczególnych regionów mózgu.
- 148 regionalnie specyficznych sQTL było skorelowanych z podwyższonym ryzykiem chorób neurodegeneracyjnych.
- Zidentyfikowano korelację między sQTL a wariantami w locus *FLOT1* związanymi z ryzykiem schizofrenii.



Liczba sQTL zidentyfikowana w poszczególnych regionach mózgu. Na niebiesko zaznaczono sQTL specyficzne dla regionu. Najwięcej sQTL zidentyfikowano w mózdzku.

Istotna korelacja między sQTL a genotypem dla genu *RTEL1*, który koduje helikazę regulującą elongację telomerów. Wariant CC jest skorelowany z giejakiem.



W analizie sQTL alternatywny egzon lub alternatywne miejsca splicingu są korelowane z genami dotyczącymi analizowanych markerów, np. SNP.

Zhang et al., 2020.

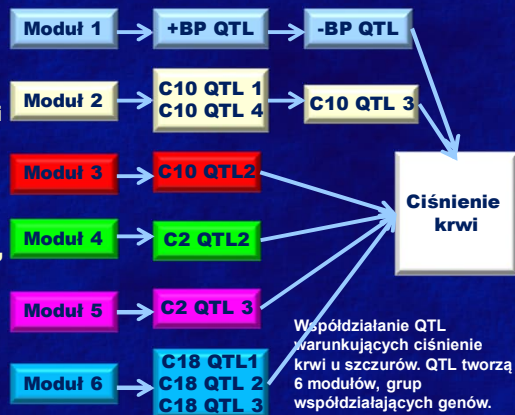


4. QTL: identyfikacja QTL

Identyfikacja genów w QTL człowieka jest utrudniona ze względu na brak możliwości manipulacji i tworzenia populacji segregujących.

Geny zidentyfikowane w QTL korelujących z ciśnieniem krwi.

- **ADD1**: koduje adducynę, białko cytoszkieletu, podnosi ciśnienie krwi przez zwiększenie zdolności absorpcji sodu przez komórki nabłonka tabularnego.
- **Cyp11B2**: koduje 11/18 beta-hydroksylazę steroidową funkcjonującą w mitochondriach, allel T zwiększa ryzyko nadciśnienia.
- Region 3p14.1-q12.3 koreluje z poziomem ciśnienia krwi w populacjach europejskich, nie zidentyfikowano genów.



Geny zlokalizowane w QTL człowieka identyfikuje się przez ekstrapolację danych uzyskanych dla organizmów modelowych, zwłaszcza ssaków.

Deng 2017

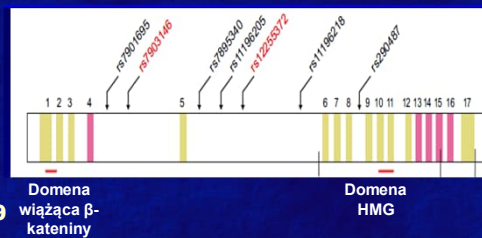


2. Mapowanie QTL: czym są QTL?

QTL w regionie 10q.25 człowieka koreluje z ryzykiem cukrzycy typu 2. Maksimum funkcji LOD odpowiada regionowi z genem **TCF7L2**.

Funkcja **TCF7L2**

- Jest to czynnik transkrypcyjny, należy do rodziny wzmocniaczy ulegających ekspresji w limfocytach T, łączy się z DNA za pomocą domeny HMG.
- Jest efektem szlaku sygnałowego Wnt obejmującego 19 glikoprotein i modulującego procesy rozwojowe i metaboliczne.
- Uczestniczy w regulacji adipogenezy, promuje rozrost tkanki tłuszczowej oraz insulinooporność.
- Reguluje geny związane z metabolizmem lipidów i glukozy.



Gen **TCF7L2** koduje czynnik transkrypcyjny, zawiera 18 egzonów. Egzony zaznaczone na żółto podlegają zawsze ekspresji, egzony różowe występują w alternatywnych transkryptach. Domena wiążąca kateniny (uczestniczą w adhezji komórek) odpowiada egzonowi 1, domena HMG (bierze udział w wiązaniu z DNA) odpowiada egzonowi 10-11. Obie domeny oraz fragment egzonu 18 od końca 3' są ewolucyjnie konserwatywne. Na górze zaznaczono polimorfizm, na czerwono - allele korelujące z ryzykiem cukrzycy typu 2.

Działanie alleli **TCF7L2** odpowiada multiplikatywnemu działaniu genów. Polimorfizm w **TCF7L2** odpowiada za 20% przypadków cukrzycy typu 2.

Bosque-Plata et al., 2021

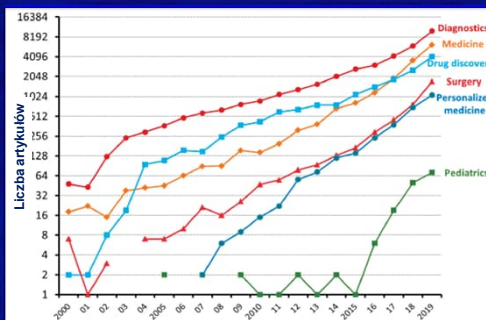


4. QTL: statystyka czy systemy uczące się

Systemy uczące się budują model matematyczny na podstawie bazy danych (zbiór uczący) w celu prognozowania.

Korzyści i ograniczenia

- Statystyka wykorzystuje ograniczoną liczbę zmiennych, wyniki są stosunkowo łatwe do interpretacji.
- Statystyka jest bardziej przydatna w typowych sytuacjach klinicznych.
- Systemy uczące się są elastyczne, nie wymagają założeń *a priori* (np. rozkład, addytywność), które nie zawsze występują w praktyce klinicznej.
- Systemy uczące się są bardziej przydatne w analizach genów, białek, ekspresji (tzw. „omiki”).



Liczba artykułów dotyczących zastosowań systemów uczących się wzrasta w latach 2000-2019. Wzrost w diagnostyce i poszukiwaniu nowych leków, a także pediatrii jest wykładniczy. Z drugiej strony brak porównania z liczbą artykułów wykorzystujących tradycyjną statystykę lub ogólną liczbą artykułów z zakresu nauk medycznych i nauk o zdrowiu.

Statystyka i systemy uczące się są komplementarne. Mogą stanowić zintegrowany system analizy danych.

Rajula et al., 2020.



Zagadnienia 1-2

1. Cechy ilościowe: definicja

- Jak można zdefiniować cechy ilościowe?
- Proszę podać przykłady cech ilościowych, w tym u człowieka.
- Które cechy podlegają najczęściej selekcji w trakcie ewolucji?
- Dlaczego analiza cech ilościowych jest trudniejsza niż cech jakościowych?

2. Cechy ilościowe: właściwości

- Ile genów koduje cechy ilościowe?
- Co to są poligeny?
- Dlaczego w przypadku cech ilościowych różnice pomiędzy fenotypami są niewielkie?
- Od czego zależy liczba klas fenotypowych w przypadku cech ilościowych?
- Proszę podać cechy rozkładu normalnego.
- Na czym polega interakcja genotypowo-środowiskowa?
- Proszę podać przykłady interakcji genotypowo-środowiskowej.
- Czy wzrost człowieka jest cechą ilościową czy jakościową? Proszę uzasadnić.
- Jak czynniki środowiskowe wpływają na wzrost człowieka?



Zagadnienia 3

3. Układ doświadczalny

- Jak można zdefiniować populację statystyczną? Czy pojęcie to jest tożsame z populacją biologiczną? Proszę uzasadnić.
- Co to jest próba?
- Czy w przypadku cech ilościowych należy analizować całą populację? Proszę uzasadnić odpowiedź.
- Czym różni się populacja skończona od nieskończonej?
- Co oznacza pojęcie badań wyczerpujących i badań cząstkowych?
- Jakie są ograniczenia związane ze stosowaniem małej próby?
- Od czego zależy minimalna wielkość próby?
- Co to jest blok i powtórzenie?
- Na czym polega układ bloków randomizowanych?
- Jak należy zaplanować doświadczenie polegające na porównaniu 5 traktowań, np. 5 różnych dawek leku?
- Czemu służy blokowanie w przypadku analizy cech ilościowych?
- Czy pojęcie randomizacji w badaniach klinicznych jest zawsze tożsame z układem bloków randomizowanych? Proszę uzasadnić.
- Na czym polega układ równoległy i krzyżowy?
- Który układ doświadczalny umożliwi porównania wewnątrzosobnicze? Proszę uzasadnić.
- Jakie warunki muszą być spełnione aby zastosować układ krzyżowy?



Zagadnienia 3-4

3. Układ doświadczalny

- Proszę wymienić elementy analizy RCT.
- Jakie punkty końcowe wykorzystuje się w analizie RCT?
- Co jest celem analizy śródkresowej w badaniach RCT?
- Na czym polegają badania kohortowe?
- Proszę wymienić ograniczenia badań kohortowych?
- Jakie najczęściej błędy popełniane są w badaniach w odniesieniu do grypy kontrolnej?
- Jak prawidłowo traktować grupę kontrolną w badaniach?
- Na czym polega różnica między układem jednoczynnikowym i wieloczynnikowym do
- Na czym polega najczęstszy błąd związany z układem doświadczenia?

4. Metody statystyczne: rozkład częstości

- Co to jest rozkład częstości?
- Proszę podać definicję histogramu.
- Kiedy obserwujemy normalny rozkład częstości w przypadku cech ilościowych?
- W jakich warunkach powstaje rozkład bimodalny? Proszę podać przykłady cech człowieka dających rozkład bimodalny.
- Kiedy powstaje rozkład asymetryczny? Proszę podać przykłady cech człowieka dających rozkład asymetryczny?



Zagadnienia 5-6

5. Metody statystyczne: miary tendencji

- Proszę zdefiniować miary tendencji centralnej? Jaką pełnią one funkcję?
- Proszę wymienić miary tendencji centralnej.
- Jak zachowują się miary tendencji centralnej w rozkładach asymetrycznych?
- Kiedy wykorzystanie średniej arytmetycznej jest uzasadnione?
- Proszę zdefiniować średnią geometryczną.
- Kiedy wykorzystujemy średnią geometryczną?



6. Metody statystyczne: miary rozproszenia

- Jaką właściwość populacji opisują miary rozproszenia?
- Proszę wymienić najczęściej stosowane miary rozproszenia.
- Co to jest rozstęp?
- Kiedy najczęściej wykorzystuje się rozstęp?
- Jak definiujemy wariancję?
- Co to jest odchylenie standardowe?
- Jeżeli średnia dwóch populacji jest taka sama, a wariancja wynosi i odpowiednio 5,0 i 0,5, to jak wygląda wykres dla analizowanych populacji?
- Co oznaczają stopnie swobody?
- Czy odchylenie standardowe i błąd standardowy to pojęcia tożsame? Proszę uzasadnić.



Zagadnienia 7-8

7. Metody statystyczne: testowanie hipotez

- Jakie jest podstawowe znaczenie analizy wariancji (ANOVA)?
- Kiedy można stosować analizę wariancji?
- W jaki sposób testuje się istotność danego efektu w analizie wariancji?
- Czy istotna wartość testu F w analizie wariancji świadczy o istnieniu różnic między wszystkimi grupami? Proszę uzasadnić.
- Czemu służy analiza posthoc?



8. QTL

- Proszę zdefiniować QTL.
- Proszę podać etapy poszukiwania genów odpowiedzialnych za uzależnienie alkoholowe.
- Co to jest ADHD i jaki jest związek tego zespołu z QTL?
- Proszę wymienić geny związane z ADHD?
- Jaką funkcję pełni gen DAT1?
- Proszę podać definicję mapowania interwałowego?
- Proszę podać elementy mapowania interwałowego?
- Co to jest interwał?
- Do czego służy funkcja LOD?
- Jaki obszar chromosomu może zajmować QTL?
- Co oznacza pojęcie „duch” QTL?



Zagadnienia 9

9. QTL, cd.

- Jakie znaczenie ma mapowanie porównawcze QTL u świni domowej i człowieka?
- Ile genów może wchodzić w skład pojedynczego QTL?
- Ile QTL odpowiada za otyłość u człowieka?
- Ile QTL odpowiada za metabolizm cholesterolu, HDL-C u człowieka?
- Proszę podać rodzaje genów, które uczestniczą w metabolizmie cholesterolu u człowieka.
- Co rozumiemy pod pojęciem eQTL?
- Czym różni się *cis* eQTL od *trans* eQTL?
- Jak SNP wpływają na poziom ekspresji genów?
- Co oznacza pojęcie sQTL?
- Czy można uzasadnić choroby neurodegeneracyjne występowaniem sQTL? Proszę uzasadnić odpowiedź.
- Jakie geny zidentyfikowano w QTL odpowiedzialnym za regulację ciśnienia krwi?
- Jak identyfikuje się geny w QTL u człowieka?
- Jaki region można powiązać z ryzykiem cukrzycy typu 2 u człowieka?
- Jaką funkcję pełni *TCF7L2*?
- Proszę omówić strukturę *TCF7L2*.




Centre for Evolution, Genomics and Biomathematics, e-Gene



prof.romanzielinski@gmail.com

<https://www.matgen.pl>

**Centre for Evolution, Genomics
and Biomathematics, e-Gene**



e-Gene

polokkornelia@gmail.com

<https://www.matgen.pl>