

Ćwiczenie 09

Metody identyfikacji cech ilościowych. Cechy ilościowe i człowieka. Metody statystyczne w analizie cech ilościowych.

Kornelia Polok

1. Metody identyfikacji cech ilościowych

1.1. Zmienność ciągła

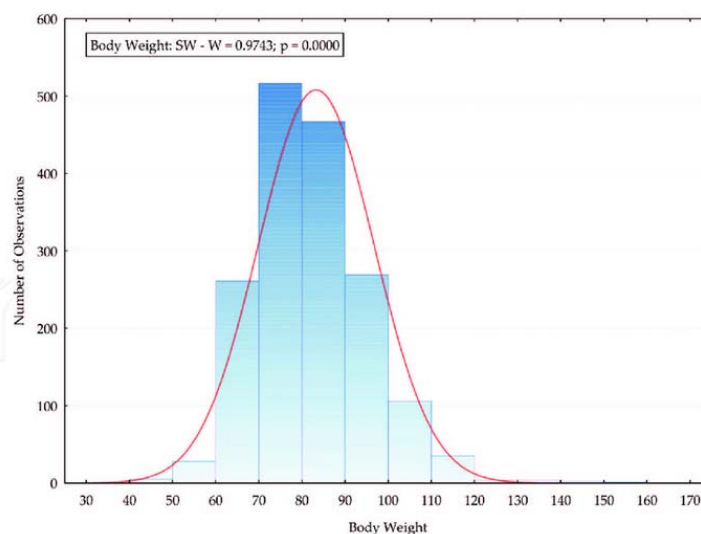
Cechy ilościowe to cechy mierzalne. Kodowane są one przez wiele genów, z których każdy ma niewielki wpływ na daną cechę. Cechy te podlegają wpływowi środowiska. Mówimy wówczas o interakcji genotypowo-środowiskowej.

➔ 1.1.1. Cechy ilościowe stanowią większość analizowanej zmienności. Podlegają one selekcji w trakcie ewolucji, a także w procesach hodowlanych. Ich analiza jest utrudniona ze względu na obecność wielu klas fenotypowych, które trudne zaklasyfikować. Dlatego w analizie cech ilościowych stosuje się metody statystyczne. Ze względu na brak wyraźnych różnic między klasami, zmienność cech ilościowych ma charakter ciągły.

1.1.2. **Zmienność ciągła** to stopniowe zmiany cechy w pewnym zakresie wartości. Przykładem jest wzrost, np. od najniższego do najwyższego, masa ciała, a także barwa, jeżeli obserwujemy cały zakres odcieni. Zmienność ciągła jest przeciwstawiana zmienności skokowej czyli zmianom nagłym, np. czerwony i biały (bez form pośrednich). W klasycznym doświadczeniu Mendla wysokość grochu była cechą zmieniającą się skokowo, gdyż obserwował on tylko formy wysokie i niskie. Takli rozkład był możliwy dzięki starannemu wyselekcjonowaniu linii do analiz. Były to linie czyste o utrwalonym wzroście. Dodatkowo różnice we wzroście były uwarunkowane zmianą w jednym genie. Podobne wyniki uzyskamy np. krzyżując karłowe i wysokie odmiany jęczmienia. U człowieka mogłaby to być osoba o wysokim wzroście i karłowata. Karłowatość związana jest z defektem w 1-2 genach. Dlatego w takich parach wzrost dziczyłby się mendlowsko. Natomiast w populacji istnieje wiele genów warunkujących wzrost i wiele alleli w każdym genie. Wzrost dziedziczy się wówczas poligenowo czyli jest

uwarunkowany wieloma genami. Zmienność wzrostu w populacji będzie miała charakter ciągły.

- Rozkład cechy, która wykazuje zmienność ciągłą jest rozkładem normalnym (Gauss lub Laplace-Gauss). Krzywa rozkładu ma kształt dzwona (Rys. 1.1.2). W rozkładzie normalnym wartości skrajne są najrzadsze, a najczęstsze są wartości średnie. Idealna krzywa rozkładu normalnego jest symetryczna czyli 50% wartości jest mniejszych od średniej i 50% wartości jest wyższych od średniej.



Rys. 1.1.2. Rozkład normalny masy ciała w populacji ludzkiej (mężczyźni).

- Wartością oczekiwaną jest wartość średnia, μ . Wykorzystuje się także medianę czyli wartość środkową oraz wartość modalną czyli wartość, która występuje najczęściej. Rozkład wartości wokół średniej określa odchylenie standardowe (σ) oraz jego kwadrat czyli wariancja (σ^2)
- Wykorzystanie rozkładu normalnego wynika z faktu, że wiele cech organizmów żywych wykazuje w populacjach taki rozkład. Ponadto, zauważono, że jeżeli cechy nie mają rozkładu normalnego, to zwiększenie liczby zmiennych często prowadzi do uzyskania rozkładu zbliżonego do normalnego.

1.2. Histogram

➔ Analiza rozkładu danej cechy w populacji wymaga sporządzenia histogramu. Histogram jest to graficzne przedstawienie częstości osobników o określonym fenotypie. Poszczególne osobniki zalicza się do klas fenotypowych, które tworzy się arbitralnie, np. co 10 cm. Liczebność osobników w danej klasie dzieli się przez liczbę wszystkich badanych osobników otrzymując częstość danej klasy fenotypowej. Czasami częstość przelicza się na procenty. Następnie częstości przedstawia się na wykresie słupkowym otrzymując histogram. Kształt histogramu pozwala na wstępną ocenę działania genów warunkujących cechy ilościowe.

- Rozkład normalny: większość genów warunkujących daną cechę ilościową wykazuje działanie addytywne (kumulatywne) co oznacza, że sumują się efekty alleli w locus oraz genów pomiędzy loci.
- Rozkład asymetryczny: występuje gdy geny współdziałają ze sobą np. występuje komplementacja, epistaza, multiplikatywne działanie genów.
- Rozkład bimodalny: na ogół świadczy o działaniu dwóch różnych czynników, występuje przy dominacji.

1.3. Sporządzanie histogramu



W tabeli poniżej przedstawiono średnie wartości dla wysokości pewnej rośliny. Na podstawie przedstawionych danych:

Proszę podzielić wyniki na klasy fenotypowe. Ile jest klas fenotypowych? Jaki przedziały podamy dla każdej klasy?(2 punkty)

Proszę podać procent osobników w każdej klasie. (2 punkty)

Na podstawie uzyskanych procentów proszę sporządzić histogram. Jaki kształt ma histogram?

Proszę podać jak mogą działać geny warunkujące analizowane cechy ilościowe. (2 punkty)

Nr rośliny	Wysokość [cm]	Nr rośliny	Wysokość [cm]
1	15	29	37
2	77	30	42
3	81	31	96
4	44	32	65
5	99	33	47
6	61	34	39
7	32	35	66
8	67	36	69
9	26	37	25
10	46	38	41
11	33	39	65
12	68	40	84
13	75	41	48
14	53	42	68
15	36	43	17
16	76	44	57
17	89	45	71
18	62	46	49
19	23	47	58
20	45	48	28
21	35	49	55
22	56	50	51
23	75	51	59
24	41	52	59
25	45	53	52
26	85	54	55
27	63	55	54
28	73	56	56

Czas wykonania: 15 minut

2. Cechy ilościowe u człowieka

2.1. Wielogenowe uwarunkowanie

➔ Cechy ilościowe często uwarunkowane są wieloma genami, których efekty sumują się zarówno w obrębie locus jak i pomiędzy poszczególnymi loci (genami). Takie działanie genów określamy jako działanie kumulatywne lub addytywne. Przykładowo mogą sumować się efekty działania locus A i B. I tak, np., homozygota aabb ma 30 cm wzrostu, a każdy allel dominujący zwiększa jej wysokość o 1 cm. Wówczas heterozygota AaBb będzie miała 32 cm, podobnie jak AAbb i aaBB, gdyż w każdym genotypie występują dwa allele dominujące. Maksymalna wysokość to 34 cm u homozygoty dominującej AABB.

2.2. Ćwiczenie



U człowieka za masę ciała odpowiadają między innymi trzy niezależne geny. Masa ciała homozygoty recesywnej aabbdd wynosi 50 kg. Każdy allel dominujący zwiększa masę ciała o 2 kg. W pewnej rodzinie matka i ojciec ważyli 56 kg a ich dorosła córka 50 kg.

- Podaj genotyp córki. (1 p)
- Podaj genotypy matki i ojca. (1p)
- Czy jest możliwe pojawienie się w tej rodzinie dziecka, które osiągnie masę 62 kg? Uzasadnij odpowiedź. Jaki genotyp będzie miało to dziecko? (2p)
- Jakie jest prawdopodobieństwo pojawienia się dziecka o genotypie podanym w punkcie D? (1p)

Czas wykonania: 15 minut

3. Metody statystyczne w analizie cech ilościowych

3.1. Analiza wariancji

➔ Analiza wariancji (ANOVA) jest podstawową metodą wykorzystywaną do porównań cech ilościowych, ale także może być wykorzystana w badaniach socjologicznych, ekonomicznych itd. Pozwala ona na podział zmienności całkowitej na zmienność uwarunkowaną genetycznie oraz środowiskowo, a także określenie interakcji genotypowo-środowiskowej.

- Analiza wariancji wymaga określenia zmiennej niezależnej, np. genotypu oraz zmiennej zależnej, np. wzrostu, który zależy od genotypu, masy ciała, dla roślin plonu. W przypadku eksperymentów mających badać wpływ jakiegoś czynnika, np. leku na stopień niszczenia patogenu, wybiera się jeden genotyp a zmienną niezależną stanowią różne dawki leku.
- W najprostszym układzie mamy do czynienia z analizą wariancji jednoczynnikową (One-way ANOVA), gdy badamy wpływ jednego czynnika na cechę, np. wysokość różnych genotypów. W tym przypadku zmienność całkowitą dzielimy na:
 - zmienność między genotypami,
 - zmienność między powtórzeniami będącą błędem doświadczenia lub zmiennością środowiskową.

- Wieloczynnikową analizę wariancji (MANOVA) stosuje się, gdy badamy wpływ wielu czynników, np. masa ciała zależy od genotypu i diety. Wówczas zmiennymi niezależnymi jest genotyp i dieta, natomiast zmienną zależną masa ciała. W tym przypadku zmienność dzielimy na:
 - ▶ zmienność genetyczną, np. różnice w masie ciała wynikające z genotypu;
 - ▶ zmienność środowiskową, np. różnice w masie wynikające z samej diety;
 - ▶ interakcję genotypowo-środowiskową, czyli jak reagują poszczególne genotypy na różnice w diecie;
 - ▶ błąd doświadczenia.
- Analiza wariancji pozwala wnioskować czy poszczególne efekty są istotne, np. czy istnieją istotne różnice pomiędzy genotypami. O istotności wnioskujemy na podstawie testu Fishera. Test Fishera to iloraz średniego kwadratu (mean square) dla genotypu (traktowania) i błędu (error). Jeżeli otrzymana wartość jest większa od wartości granicznej to efekt genotypu jest istotny.
- W programach statystycznych często jest podana wartość F oraz wartość P (prawdopodobieństwo) dla tego F. Istotność odczytujemy korzystając z wartości P. Przyjmuje się, że różnice są istotne statystycznie jeżeli P jest mniejsze lub równe 0,05. Oznacza to, że prawdopodobieństwo iż otrzymany wynik nie jest efektem istotnych różnic jest niskie i wynosi 5%. Innymi słowy, te 5% oznacza, że różnice mają charakter losowy, przypadkowy.
 - ▶ **Dla hipotezy zerowej, że różnice pomiędzy genotypami są istotne statystycznie, jeżeli wartość $F = 0,001$ to oznacza, że prawdopodobieństwo iż obserwowana zmienność nie jest efektem istotnych różnic wynosi 0,1%. Ponieważ jest mniejsza od 5% to przyjmujemy hipotezę zerową.**
 - ▶ **Dla hipotezy zerowej, że różnice pomiędzy genotypami są istotne statystycznie, jeżeli $F = 0,76$ to oznacza, że prawdopodobieństwo iż obserwowana zmienność nie jest efektem istotnych różnic statystycznych wynosi 76%. Jest wysokie i hipotezę zerową należy odrzucić. Badane genotypy nie różnią się.**
- W praktyce obliczenia wykonuje się w programach statystycznych i całość obliczeń sprowadza się do umiejętności wprowadzenia danych oraz ich interpretacji.
- Analiza wariancji pozwala stwierdzić czy istnieją różnice między genotypami, ale nie określa czy różnice dotyczą każdej pary genotypów oddzielnie. W tym celu stosuje się testy istotności tzw. Post-hoc jak test LSD, Tukey'a, t-studenta i inne.

3.2. Przykład obliczeń

Zbadano wzrost mieszkańców 4 regionów Polski oznaczonych jako X1, X2, X3, i X4. Jednocześnie zbadano haplotypy osób uczestniczących w eksperymencie i okazało się, że istnieją różnice genotypowe w locus *Ad1*. W populacji X1 i X2 najczęściej występował allel A1, w populacji X3 allel A2 natomiast w populacji X4 allel A3. Poniżej przedstawiono tabelę ze wzrostem badanych osób w centymetrach.

Populacja	X1	X2	X3	X4
Wzrost	156	160	175	175
	160	159	176	165
	161	158	178	181
	157	161	169	156
	158	155	165	160
	160	158	170	178
	158	160	172	172
	161	162	176	169
Średnia	158.9^a	159.1^a	172,6^b	169,5^b
Odchylenie standardowe	1,9	2,2	4,4	8,7

3.2.1. Wynik analizy wariancji

Hover over the cells for more information.

Source	DF	Sum of Square	Mean Square	F Statistic	P-value
Groups (between groups)	3	1203.343750	401.114583	15.435436	0.00000406027
Error (within groups)	28	727.624931	25.986605		
Total	31	1930.968681	62.289312		

R code.

- Wartość statystyki F dla różnic pomiędzy populacjami: **15.4354**
- Wartość P dla różnic między populacjami: **P = 0.00000406027 = 4,1 x 10⁻⁶**
- Różnice między populacjami są istotne statystycznie, ale ta analiza nie podaje, które populacje się różnią, a które nie. Wiemy tylko, że ogólnie

3.2.2. Test Tukey'a pozwala określić różnice między poszczególnymi populacjami.

Group	x2	x3	x4
x1	0.25	13.75	10.63
x2	0.0	13.50	10.38
x3	13.50	0.0	3.13

- Na podstawie testu Tukey'a populacje, które mają ten sam wzrost to:
 - ▶ X1 i X2
 - ▶ X3 i X4
- Populacje, które mają różny wzrost to:
 - ▶ X1 i X3
 - ▶ X1 i X4
 - ▶ X2 i X3
 - ▶ X2 i X4
- Na podstawie uzyskanych wyników, możemy uszeregować populację od najwyższych do najniższych zaznaczając istotność.
 - ▶ X3: 172,6 a
 - ▶ X4: 169,5 a
 - ▶ X2: 159,1 b
 - ▶ X1: 158,9 b

Odpowiedzi

1. Metody identyfikacji cech ilościowych

1.3. Sporządzanie histogramu

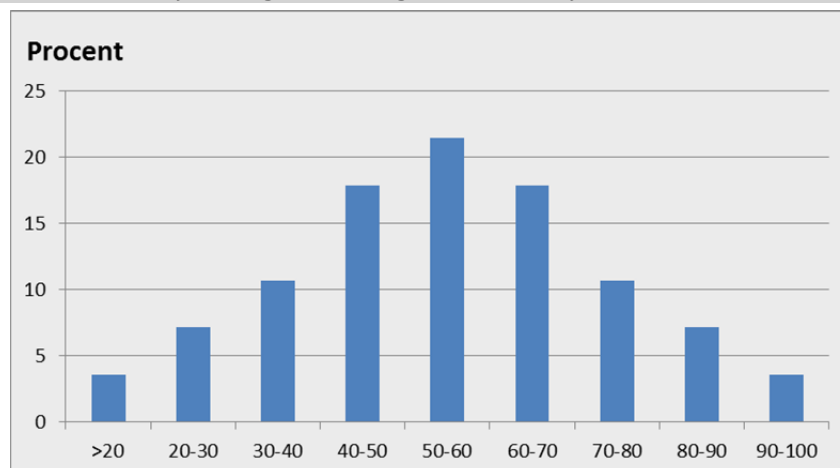
A. Proszę podzielić wyniki na klasy fenotypowe. Ile jest klas fenotypowych? Jaki przedziały podamy dla każdej klasy?(2 punkty)

Dzielimy na klasy co 10 cm: 10-20, 20-30 itd. Klas jest 9.

Proszę podać procent osobników w każdej klasie. (2 punkty)

Przedziały	Liczebność	Procent
>20	2	4
20-30	4	7
30-40	6	11
40-50	10	18
50-60	12	21
60-70	10	18
70-80	6	11
80-90	4	7
90-100	2	4
Suma	56	

Na podstawie uzyskanych procentów proszę sporządzić histogram. Jaki kształt ma histogram? Proszę podać jak mogą działać geny warunkujące analizowane cechy ilościowe. (2 punkty)



Uzyskany wykres jest zgodny z rozkładem normalnym (krzywa Gaussa), co świadczy o działaniu addydywnym (kumulatywnym) większości genów odpowiedzialnych za wzrost.

2. Cechy ilościowe u człowieka

2.2. Zadanie

U człowieka za masę ciała odpowiadają między innymi trzy niezależne geny. Masa ciała homozygoty recesywnej aabbdd wynosi 50 kg. Każdy allel dominujący zwiększa masę ciała o 2 kg. W pewnej rodzinie matka i ojciec ważyli 56 kg a ich dorosła córka 50 kg.

A. Podaj genotyp córki. (1 p)

Córka ważyła 50 kg, a więc miała masę typową dla homozygoty recesywnej względem wszystkich trzech genów, zatem genotyp córki był: **aabbdd**

B. Podaj genotypy matki i ojca. (1p)

Rodzice ważyli 56 kg każdy, a więc musieli mieć 3 allele dominujące. Ze względu na masę mogły one być w dowolnym układzie np. AABbdd, AaBBdd etc. Jednakże w rodzinie tej urodziła się córka, która była homozygotą recesywną. Oznacza to, że każdy z rodziców musiał mieć allel recesywny w każdym z 3 genów. W tej sytuacji jedyną możliwością istnienia 3 alleli dominujących u każdego z rodziców jest układ heterozygotyczny zatem:

- ▶ matka: **AaBbDd**
- ▶ ojciec **AaBbDd**

C. Czy jest możliwe pojawienie się w tej rodzinie dziecka, które osiągnie masę 62 kg? Uzasadnij odpowiedź. Jaki genotyp będzie miało to dziecko?(2p)

- ▶ Tak
- ▶ Jest to krzyżowanie 2 heterozygot względem 3 genów. Heterozygota taka wytwarza 8 typów gamet, między innymi gametę ABD. Połączenie gamety od matki ABD i gamety ABD od ojca da genotyp AABBDD.
- ▶ Masa AABBDD wyniesie 50 kg plus 6 x 2 kg na każdy allel dominujący czyli
- ▶ $50 \text{ kg} + 2 \text{ kg} \times 6 = 62 \text{ kg}$
- ▶ **AABBDD**

D. Jakie jest prawdopodobieństwo pojawienia się dziecka o genotypie podanym w punkcie D? (1p)

- ▶ 1/64
- ▶ Każda z heterozygot wytworzy 8 gamet. Oznacza to, że liczba możliwych kombinacji wyniesie $8 \times 8 = 64$ genotypy.
- ▶ Genotyp AABBDD powstanie tylko w jednym przypadku, z połączenia dwóch gamet ABD.