

J. PERKAL (Wrocław)

PROJEKT PEWNEGO TESTU GENETYCZNEGO

1. Wstęp. Prof. Lush (Iowa University) podczas seminarium genetycznego w Łopusznej (czerwiec-lipiec 1960) zwrócił uwagę uczestników seminarium na następujące zagadnienie (patrz [1]):

Zwierzęta hodowlane mogą zawdzięczać swoją wysoką wartość (użytkową) współdziałaniu dwu różnych genów tego samego lokusa, albo różnych lokusów. W pierwszej z tych sytuacji, tzw. *over-dominacji* (patrz [3], str. 362), heterozygotyczny zespół genów np. Aa ma większą wartość genotypową niż zespoły homozygotyczne: AA lub aa. Druga sytuacja zwana *epistazą* oznacza, że współdziałanie np. zespołu AA w jednym lokusie z zespołem BB w drugim lokusie daje większą wartość genotypową niż współdziałanie jakichkolwiek innych zespołów genów w tych dwóch lokusach.

W przypadku *over-dominacji* hodowca powinien wyselekcjonować dwie czyste linie zwierząt hodowlanych, tj. linię o zespole genów AA oraz linię o zespole aa w wyróżnionym lokusie. Krzyżując te dwie czyste linie hodowca otrzyma pierwsze pokolenie mieszańców o zespole genów Aa w wyróżnionym lokusie, czyli same takie osobniki, które jako heterozygotyczne będą miały, pod wpływem wykrytej *over-dominacji*, najwyższą wartość genotypową. Wytwarzanie takiego pierwszego pokolenia mieszańców, zwanego też krzyżówką produkcyjną, jest szeroko stosowane w krajach o wysokim poziomie hodowli. Warto przypomnieć, że następne pokolenia mieszańców zawierają już prócz zespołów heterozygotycznych również homozygotyczne w wyróżnionym lokusie, powstałe wskutek rozszczepiania się heterozygot (mieszańców).

W przypadku *epistazy* hodowca może wyselekcjonować jedną linię o pożądanym zespole genów w dwóch wyróżnionych lokusach. Można tak postąpić, ponieważ w każdym z tych lokusów pożądanym jest zespół homozygotyczny: w jednym lokusie AA, w drugim BB. Wyhodowana linia będzie stabilna ze względu na własności epistatyczne, gdyż efekt epistazy, związany z występowaniem zespołu AA w pierwszym lokusie i zespołu BB w drugim — będzie dziedziczny. Poprawę wartości użytkowej

można w tym przypadku (epistazy) osiągnąć zwykłymi metodami selekcji i utrzymywać przez niedopuszczanie obcej krwi.

W praktyce, gdy hodowca obserwuje np. fermę kur, może się szybko zorientować, które kury znoszą dużo jaj (są dobre), a które mało (złe). Jednakże trudno stwierdzić, czy u dobrych kur duża nieśność wynika z over-dominacji, czy z epistazy (z współdziałania genów tego samego, czy różnych lokusów). Czasem można takie rozpoznanie osiągnąć na podstawie ksiąg rodowodowych z obserwacjami dużej ilości przodków. Na ogół jednak nie umiemy rozpoznać z którą sytuacją mamy do czynienia. Nie ma testu genetycznego do rozróżniania over-dominacji od epistazy. O postępowaniu hodowlanym często decyduje intuicja lub przypadek. Nie trzeba chyba zwracać uwagi czytelnika na olbrzymie znaczenie teoretyczne i praktyczne (w hodowli) umiejętności rozróżniania over-dominacji od epistazy.

Prof. Lush zwracając się do uczestników seminarium w Łopusznej, a w szczególności do matematyków, zaproponował zajęcie się tym zagadnieniem. W pracy niniejszej proponuję test probabilistyczno-genetyczny do odróżniania over-dominacji od epistazy. Test ten ma działać w sytuacji genetycznie uproszczonej (rozważam tylko 1 lub 2 lokusy). Wyślę go i opiszę dla przypadku kur, chociaż powinien on działać i w innych hodowlach.

Test i przykłady

2. Test. Wybieramy wszystkie dobre kury mające co najmniej po 5 córek o sprawdzonej nieśności. Liczbę tych kur oznaczamy przez n . Przez *potomstwo* każdej z nich rozumiemy jej 5 córek (w przypadku, gdy kura ma więcej niż 5 córek, potomstwo wybieramy losowo). Spośród wszystkich $5n$ córek wybieramy połowę $\left(\frac{5n}{2}\right)$ o największej nieśności i te córki nazywamy *dobrymi*. Przez t_i (gdzie $i = 1, 2, \dots, n$) oznaczamy liczbę dobrych kur w potomstwie i -tej kury. Jeśli wariancja empiryczna liczb t_i jest normalna⁽¹⁾ czyli nie istotnie różna od 1,25, mamy do czynienia z over-dominacją. Jeśli ta wariancja jest nadnormalna czyli większa od 1,25 — mamy do czynienia z epistazą. (Dokładniejszą analizę sytuacji, w przypadku, gdy wariancja jest nie istotnie nadnormalna, opisuje w ustępie 7.)

3. Przykłady. A. W fermie drobiu w Czechnicy (dane z ksiąg z r. 1954) pewną ilość kur uznano za dobre i włączono do stadek selekcyjnych.

⁽¹⁾ Wariancję empiryczną nazwiemy tu *normalną*, jeśli równa się ona wariancji teoretycznej $\sigma^2 = kp(1-p)$ zmiennej losowej o rozkładzie binomialnym z parametrami k i p .

10 spośród tych kur miało potomstwa (po co najmniej 5 córek). W przypadku, gdy kura miała więcej niż 5 córek wybraliśmy z nich losowo potomstwa. Dwie z kur miały po 2 potomstwa (po 10 córek). W ten sposób wybraliśmy 12 potomstw, tj. $12 \times 5 = 60$ córek. Medianą ich rocznej nieśności okazała się liczba 182 jaj. A więc za dobre córki uznajemy te, których nieśność wynosi co najmniej 182 jaja rocznie.

W tabelicy I podano liczby t_i dobrych kur w poszczególnych potomstwach:

TABLICA I

Liczby t_i dobrych kur w poszczególnych potomstwach ($i = 1, 2, \dots, 12$)

i	1	2	3	4	5	6	7	8	9	10	11	12
t_i	0	0	1	1	3	3	3	3	3	4	4	5

Liczby tej tabelicy można zestawzić w szereg rozdzielczy mówiący ile razy w tabelicy I powtarzają się poszczególne wartości liczby t .

TABLICA II

Szereg rozdzielczy liczby t dobrych kur w $n = 12$ potomstwach

Liczba dobrych kur t	0	1	2	3	4	5	Razem
Liczba n_t potomstw o t dobrych kurach	2	2	0	5	2	1	12

W znany sposób obliczamy z tego szeregu rozdzielczego średnią arytmetyczną \bar{t} oraz empiryczną wariancję s_t^2 (nie obciążoną błędem systematycznym).

$$\bar{t} = \frac{1}{n} \sum t n_t = \frac{30}{12} = 2,50,$$

$$s_t^2 = \frac{n}{n-1} (\bar{t}^2 - \bar{t}^2) = \frac{12}{11} (8,67 - 6,25) = 2,64.$$

Porównując wariancję empiryczną $s_t^2 = 2,64$ z teoretyczną $\sigma_t^2 = 1,25$ widzimy, że wariancja empiryczna jest nadnormalna. A więc w myśl naszego testu dobroć kur jest w tej fermie wynikiem epistazy.

B. W tej samej fermie 4 lata później (dane z ksiąg z r. 1958), w taki sam sposób jak w poprzednim przykładzie, wybrano 20 potomstw (18 kur). Medianą nieśności $5 \times 20 = 100$ córek okazało się 196 jaj rocznie. Tablica III przedstawia szereg rozdzielczy liczb t : dobrych córek (co najmniej 196 jaj) w tych 20 potomstwach.

TABLICA III

t	0	1	2	3	4	5	Razem
n_t	2	1	6	7	3	1	20

W taki sam sposób jak wyżej obliczamy $\bar{t} = 2,55$ oraz $s_t^2 = 1,63$. Jak widzimy, wariancja empiryczna jest nieco większa od teoretycznej (1,25), lecz znacznie mniejsza niż przed czterema laty. Jeśli więc występuje w tym przypadku efekt epistazy, to w każdym razie znacznie słabiej niż w poprzednim (w 1954 roku).

C. Z materiałów stacji selekcyjnej w Chorzelowie (księgi z r. 1958) wybraliśmy 33 potomstwa, czyli $33 \times 5 = 165$ kur. Medianą nieśności okazała się nieśność 149 jaj rocznie. Kury o większej nieśności nazwaliśmy dobrymi. Tablica IV przedstawia szereg rozdzielnicy zmiennej t , czyli liczby dobrych kur w potomstwie:

TABLICA IV

t	0	1	2	3	4	5	Razem
n_t	1	5	10	12	4	1	33

Średnią arytmetyczną jest $\bar{t} = 2,55$, a wariancją empiryczną $s_t^2 = 1,20$. Jest to wariancja normalna, gdyż tylko nieznacznie mniejsza od teoretycznej wariancji 1,25. To, że dobre kury tej stacji znoszą więcej jaj niż pozostałe, należy przypisać over-dominacji.

Uzasadnienie testu

4. Założenia. Zakładamy, że wyróżniona cecha (np. nieśność kur) zwierząt hodowlanych zależy od dwóch zasadniczych czynników: od właściwości genetycznych i od środowiska. Zajmiemy się tylko pierwszym z nich. Zakładamy dalej, że wyróżniona cecha zależy od jednego lub od co najwyżej dwóch lokusów, przy czym lokus mieści 2 geny wybrane z allela odpowiadającego lokusowi i zawierającego geny dwóch rodzajów. Pierwszy lokus nazwiemy literą α , a geny w odpowiednim allelu oznaczymy przez A oraz a. W lokusie α mogą wystąpić 3 zespoły genów: homozygotyczne AA oraz aa i heterozygotyczny Aa. Analogicznie drugi lokus oznaczymy przez β , a geny w odpowiednim allelu przez B i b. W tym lokusie możliwe są następujące 3 zespoły genów: BB, bb i Bb.

Rozpatrzmy najpierw przypadki, w których wyróżniona cecha zależy od jednego lokusa α . W tym przypadku, jeśli podzielimy zwierzęta według zespołu genów w lokusie α (abstrahujemy od praktycznej wykonalności tego podziału), to okaże się, że najwyższą wartość wyróżnionej cechy mają osobniki o zespole genów AA, albo Aa, albo jedne i drugie (gdyby najwyższą wartość dawał zespół genów aa można by gen a oznaczyć przez A). W pierwszej i trzeciej sytuacji homozygotyczne zespoły genów AA dają najwyższą wartość wyróżnionej cechy. Można wyselekcjonować populację o tym homozygotycznym zespole genów AA w lokusie α . Taka populacja będzie genetycznie stabilna (nie uwzględniamy tu mu-

tacji) i optymalna ze względu na wyróżnioną cechę. W drugiej sytuacji najwyższą wartość wyróżnionej cechy daje heterozygotyczny zespół genów Aa w lokusie α . O tej sytuacji, zwanej over-dominacją, będziemy pisali niżej.

Jeśli wyróżniona cecha zależy od obu lokusów α i β , możliwych jest 9 różnych zespołów genów w tych lokusach, a mianowicie: 1. $AABB$, 2. $AABb$, 3. $AAbb$, 4. $AaBB$, 5. $AaBb$, 6. $Aabb$, 7. $aaBB$, 8. $aaBb$ i 9. $aabb$. (Ze względów poligraficznych zespoły te będą oznaczał cyframi.) Podobnie jak wyżej, można określić, co rozumiemy przez zespół genów o najwyższej wartości wyróżnionej cechy. Epistazą nazywamy taką sytuację, że zespół 1 ($AABB$) daje najwyższą wartość tej cechy. Przez prostą zamianę liter A i a lub B i b widzimy, że jeśli najwyższą wartość wyróżnionej cechy daje zespół 3, 7 lub 9 mamy również epistazę. Może się zdarzyć, że najwyższą wartość wyróżnionej cechy daje zespół genów homozygotyczny w jednym, a heterozygotyczny w drugim lokusie: 2, 4, 6 lub 8. W takiej sytuacji można (przynajmniej teoretycznie) wyselekcjonować populację tak, aby otrzymać stabilny układ homozygotyczny w jednym z lokusów. W tej populacji wyróżniona cecha będzie zależała tylko od pozostałego, tj. jednego lokusa i będzie to over-dominacja. Wreszcie możliwe jest, że najwyższą wartość cechy daje zespół 5 ($AaBb$) podwójnie heterozygotyczny. Jeśli lokusy α i β są silnie skorelowane (statystycznie i genetycznie), można tę sytuację traktować jako over-dominację, w innych przypadkach — nie. Tej interesującej sytuacji nie będziemy rozważać w niniejszej pracy.

Zakładamy, zgodnie z zasadami genetyki, że jeden z genów zespołu potomka dziedziczy się losowo po matce, a drugi również losowo po ojcu, niezależnie od pierwszego. Zakładamy też, że każdy potomek dziedziczy zespół genów niezależnie od innych potomków.

Wysłowimy teraz 2 konkurencyjne hipotezy H_o (hipoteza over-dominacji) i H_e (hipoteza epistazy):

H_o : Spośród kur o zespołach genów AA , Aa i aa na wyróżnionym lokusie, największą nieśność mają kury o zespole genów Aa .

H_e : Spośród kur o zespołach genów: 1, 2, ..., 9 na wyróżnionych lokusach największą nieśność mają kury o zespole genów

- a. 1 ($AABB$),
- b. 1 i 2 ($AABB$ i $AABb$, dominacja genu B),
- c. 1 i 4 ($AABB$ i $AaBB$, dominacja genu A),
- d. 1, 2, 4 i 5 ($AABB$, $AABb$, $AaBB$ i $AaBb$, dominacja genów A i B)

Uwaga: Hipoteza H_e jest spełniona, jeśli spełniona jest jedna z 4 jej wersji a, b, c lub d.

5. Over-dominacja. Zbadajmy konsekwencje założenia H_0 . Jakie może być w tym przypadku potomstwo dobrej kury i w szczególności, jakie jest prawdopodobieństwo, żeby córka dobrej kury była też dobra? W pierwszym wierszu tablicy V podane są wszelkie możliwe zespoły genów koguta, który wraz z dobrą kurą, czyli z kurą o zespole Aa, daje badane potomstwo. W dalszych wierszach tej tablicy podane są różne możliwe zespoły genów córek i prawdopodobieństwa pojawienia się tych zespołów (w nawiasach). W ostatnim wierszu podane są prawdopodobieństwa tego, że córka dobrej kury i koguta o zespole genów zapisanym w pierwszym wierszu — będzie dobra.

TABLICA V

Zespoły genów

Kury \ Koguta	AA	Aa	aa
	AA (0,50) Aa (0,50)	AA (0,25) Aa (0,50) aa (0,25)	Aa (0,50) aa (0,50)
Aa			
Prawdopodobieństwo dobrej córki (Aa)	0,50	0,50	0,50

Widzimy więc, że bez względu na to, z jakim kogutem skrzyżowano dobrą kurę, prawdopodobieństwo tego, że córka z tej krzyżówki będzie dobra, wynosi 0,50. Nazwijmy *potomstwem* k córek dobrej kury. Wobec losowości i niezależności dziedziczenia, dobre córki w potomstwie będą się spotykały tak, jak białe kule w serii k kul wybranych losowo z urny, w której białe kule stanowią połowę (a resztę, czyli też połowę, stanowią kule innego koloru). Jak wiadomo, liczba t białych kul w takiej serii jest zmienną losową o rozkładzie binomialnym z parametrami $p = 0,50$ i k . Oczekiwaną wartością tej zmiennej losowej jest $Et = kp = 0,5k$, a wariancją (binomialną, czyli Bernoulliego) jest $\sigma^2 = kpq = 0,25k$. Stąd, że oczekiwana liczba dobrych córek w potomstwie (liczącym k córek) wynosi $0,5k$, wynika sposób określenia dobrej córki. Mianowicie, z dostępnego nam materiału badanej fermy notujemy nieśność każdej z córek dobrych matek. Znajdujemy medianę m tych nieśności, tj. liczbę dzielącą nieśności na dwie równoliczne części: nieśności nie większe od m i nieśności większe od m . Córki o nieśności większej od mediany m uznamy za dobre.

Z założenia H_0 wynika więc, że (jeśli połowę córek o lepszej nieśności uznamy za dobre córki) wariancja empiryczna liczby dobrych córek w potomstwie dobrej kury (liczącym k córek) powinna wynosić $0,25k$. Dokładniej, różnica między empiryczną wariancją s^2 a teoretyczną $0,25k$ nie powinna być istotna. Zmienna losowa t ma rozkład binomialny, więc w przybliżeniu, normalny; istotność różnicy $s^2 - 0,25k$ można ocenić za

pomoć kryterium F Snedecora (zob. np. [2], str. 21-22 i 82-87). W tablicy VI podane są przedziały wariancji empirycznej dla kilku ilości potomstw i kilku poziomów istotności; są one obliczone dla $k = 5$ córek w potomstwie.

TABLICA VI

Przedziały dla wariancji empirycznej obliczonej przy $k = 5$, dla liczby n potomstw i poziomu istotności ε

$\varepsilon \backslash n$	12	20	33	50	100
0,25	0,86-1,56	0,96-1,50	1,03-1,45	1,07-1,41	1,12-1,37
0,10	0,63-1,97	0,77-1,79	0,87-1,66	0,94-1,58	1,03-1,48
0,05	0,52-2,24	0,66-1,98	0,79-1,80	0,87-1,69	0,98-1,55
0,01	0,35-2,81	0,50-2,38	0,64-2,09	0,74-1,90	0,87-1,70

W przykładzie A otrzymaliśmy, dla $n = 12$ potomstw, wariancję empiryczną $s_e^2 = 2,64$. Jak widzimy, nie mieści się ona w przedziałach pierwszej kolumny dla $\varepsilon = 0,25$, $\varepsilon = 0,10$ i $\varepsilon = 0,05$ tablicy VI, ale mieści się ona w przedziale dla $\varepsilon = 0,01$. Wynika stąd, że nadnormalność tej wariancji jest niemal pewna. Dokładniej mówiąc, wiarygodność hipotezy twierdzącej, że wariancja empiryczna jest normalna (różni się od normalnej tylko losowo) jest mała, bo co prawda większa od 1%, lecz mniejsza od 5%.

6. Epistaza. Zbadamy konsekwencje założenia H_e , policzymy więc jakie może być potomstwo dobrej kury i z jakim prawdopodobieństwem córka dobrej kury będzie dobrą. Tablica VII jest analogiczna do tablicy V. W pierwszym jej wierszu podane są różne zespoły genów koguta, w pierwszej kolumnie różne zespoły genów kur-matek (więc tylko dobrych), a w polach na skrzyżowaniu kolumny z wierszem podane są zespoły genów, jakie córka mogła odziedziczyć po odpowiednich rodzicach.

TABLICA VII

Zespoły genów koguta (w pierwszym wierszu), kury (w pierwszej kolumnie) i córek (na polach tablicy)

$\text{♀} \backslash \text{♂}$	1	2	3	4	5	6	7	8	9
1	1	1,2	2	1,4	1,2 4,5	2,5	4	4,5	5
2	1,2	1,2,3	2,3	1,2 4,5	1,2,3 4,5,6	2,3 5,6	4,5	4,5,6	5,6
4	1,4	1,2 4,5	2,5	1,4,7	1,2,4 5,7,8 1,2,3	2,5,8	4,7	4,5 7,8	5,8
5	1,2 4,5	1,2,3 4,5,6	2,3 4,5	1,2,4 5,7,8	4,5,6 7,8,9	2,3,5 6,8,9	4,5 7,8	4,5,6 7,8,9	5,6 8,9

W tablicy VIII podane są prawdopodobieństwa tego, że córka odpowiedniej pary rodziców okaże się dobrą.

TABLICA VIII

Prawdopodobieństwa, że córki dobrych kur i różnych kogutów są dobre

Przy hipotezie H_e , a:

σ	1	2	3	4	5	6	7	8	9
1	1,00	0,50	0,00	0,50	0,25	0,00	0,00	0,00	0,00

Przy hipotezie H_e^1 , b:

σ	1	2	3	4	5	6	7	8	9
1	1,00	1,00	1,00	0,50	0,50	0,50	0,00	0,00	0,00
2	1,00	0,75	0,50	0,50	0,375	0,25	0,00	0,00	0,00

Przy hipotezie H_e , c:

σ	1	2	3	4	5	6	7	8	9
1	1,00	0,50	0,00	1,00	0,50	0,00	1,00	0,50	0,00
4	1,00	0,50	0,00	0,75	0,375	0,00	0,50	0,25	0,00

Przy hipotezie H_e , d:

σ	1	2	3	4	5	6	7	8	9
1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	1,00	0,75	0,50	1,00	0,75	0,50	1,00	0,75	0,50
4	1,00	1,00	1,00	0,75	0,75	0,75	0,50	0,50	0,50
5	1,00	0,75	0,50	0,75	0,5625	0,375	0,50	0,375	0,50

Z tablicy VIII widzimy, że każde z założeń o epistazie prowadzi do tego, iż prawdopodobieństwo dobrej córki jest zmienne, zależne od koguta i w przypadkach b, c, d zależne również od kury-matki. A więc sytuacja jest całkiem inna niż w przypadku over-dominacji, gdyż tam prawdopodobieństwo dobrej córki było stale równe 0,5, niezależnie od koguta. Nie możemy już w przypadku epistazy stosować schematu urny z kulami o stałej frakcji kul białych. Trzeba teraz zastosować inny schemat urnowy: Mamy mianowicie wiele urn, np. w przypadku hipotezy H_e , a mamy 9 urn; w pierwszej białe kule stanowią 100%, w następnych dwóch — 50%, w jednej 25% i w pozostałych 0%. Wybieramy na chybił trafił jedną z urn, a z niej wybieramy ze zwracaniem k ($= 5$) kul. Przez zmienną losową t rozumiemy liczbę białych kul wśród wybranych k kul. Analizę tego zagadnienia zajmował się H. Steinhaus i udowodnił w pracy [4], że wariancja zmiennej losowej t jest nadnormalna. Schemat urnowy

o zmiennych prawdopodobieństwach kul białych w poszczególnych urnach nazywa się *schematem Lexisa* (patrz np. [2]).

Można twierdzić, że 9 urn wspomnianych wyżej tytułem przykładu nie stanowi dobrego modelu dla naszej sytuacji, gdyż koguty o 9 różnych zespołach genów nie są na fermie hodowlanej równie prawdopodobne. Zauważmy jednak, że nie bardzo wiadomo, co to są koguty dobre, gdyż już przy hipotezie b lub c koguty o zespołach 3 lub 7 (więc o zespołach znamienych dla nie dobrych kur) dają z dobrymi kurami o pewnych zespołach dobre potomstwo w 100%. Aby uwzględnić różne prawdopodobieństwa różnych kogutów należy na każdej urnie (symbolizującej koguta i kurę) zapisać prawdopodobieństwo spotkania się takiej pary na fermie i losować urny nie jako równie często występujące, lecz z uwzględnieniem zapisanych prawdopodobieństw. Nie zmieni to jednak faktu, że prawdopodobieństwo białej kuli będzie się zmieniało od serii do serii, czyli, że wariancja zmiennej t będzie nadnormalna.

7. Porównanie wiarogodności hipotez H_o i H_e . Z założenia over-dominacji wynikało, że oczekiwana wartość zmiennej losowej t wynosi 2,5, więc wariancja binomialna 1,25. Z założenia epistazy nie możemy wysnuć podobnych wniosków. Prawdopodobieństwo p tego, że córka dobrej kury będzie dobra, zmienia się w szerokich granicach od 0 do 1 w zależności od zespołu genów dobrej matki i dowolnego ojca. W przypadku over-dominacji wiedzieliśmy, że połowę córek należy uznać za dobre. Przy założeniu epistazy nie wiemy jaką część córek należy uznać za dobre, nie wiemy więc jaką wariancję uznać za teoretyczną, tj. tę z którą należy porównywać wariancję empiryczną w celu sprawdzenia jej nadnormalności.

Hipoteza over-dominacji pozwalała wysnuć jedną konkretną wartość 0,25k dla wariancji rozważanej zmiennej losowej t . Taką hipotezę statystycy nazywają *hipotezą prostą*. Natomiast hipoteza epistazy nie pozwala wysnuć podobnego wniosku o wariancji. Wiemy tylko, że jest ona nadnormalna, czyli musi spełniać pewną nierówność. Taką hipotezę, według której parametr (w naszym przypadku wariancja) ma przyjmować nie jedną konkretną wartość, lecz dowolną wartość z jakiegoś zbioru wartości, np. z przedziału — statystycy nazywają *hipotezą złożoną*. Hipoteza epistazy jest więc hipotezą złożoną.

Zadanie, jakie stoi przed nami, polega na tym, żeby na podstawie zaobserwowanych wartości zmiennej losowej t orzec, którą z dwóch hipotez H_o czy H_e należy uznać za prawdziwą. W przypadku gdyby obie hipotezy były proste powiedzielibyśmy, że należy uznać za prawdziwą tę hipotezę, której wiarogodność jest większa. Jednakże taki sposób porównywania dwóch hipotez, z których jedna jest prosta, a druga złożona, nie jest możliwy. W praktyce w podobnych przypadkach zastępujemy zazwyczaj hipotezę złożoną przez hipotezę prostą H_1 , zawartą w hi-

potezie złożonej H_e i najbliższą hipotezie konkurencyjnej H_o . W naszym przypadku należałoby z hipotezy złożonej H_e , mówiącej, że wariancja jest większa od binomialnej, wybrać hipotezę H_1 najbliższą hipotezie H_o , mówiącej, że wariancja ta jest równa binomialnej. Ale wybranie takiej hipotezy H_1 jest niemożliwe, gdyż nie istnieje najmniejsza z liczb rzeczywistych większych od wariancji binomialnej (ani od dowolnej konkretnej liczby rzeczywistej). Zawodzi więc w naszym przypadku i ten sposób. Pozostaje analiza logiczna.

Przyjmijmy zgodnie z hipotezą H_o połowę córek (o lepszej nieśności) za dobre. Jeśli wariancja empiryczna zmiennej losowej t nie przekracza wariancji binomialnej (1,25) i zawiera się w odpowiednim przedziale podanym w tabeli VI (przy poziomie istotności 0,05 lub 0,01), to hipotezę H_o należy uznać za wiarygodniejszą od hipotezy H_e i uznać, że mamy do czynienia z over-dominacją. Jeśli w tym przypadku sprawdzimy jeszcze, że i przy innym ustaleniu frakcji dobrych kur (mniej niż 0,5 lub więcej niż 0,5) wariancja liczby dobrych kur w potomstwie nie różni się istotnie od binomialnej — utwierdzi nas to w przekonaniu o słuszności naszej decyzji.

Gdyby wariancja empiryczna była istotnie podnormalna, obie hipotezy H_o i H_e uznałibyśmy za fałszywe.

Gdyby się okazało, że wariancja empiryczna zmiennej losowej t wykacza na prawo poza odpowiedni przedział z tabeli VI przy poziomie istotności ε (np. 0,25), powinniśmy stwierdzić, że wiarygodność hipotezy H_o jest mała, a spośród hipotez prostych należących do H_e można wybrać taką, której wiarygodność jest 1 (mianowicie hipotezę mówiącą, że wariancja epistatyczna równa się empirycznej). W takim więc przypadku należy przyjąć hipotezę H_e .

Pozostaje do rozstrzygnięcia przypadek, gdy wariancja empiryczna jest zawarta między binomialną, a prawym krańcem przedziału podanego w tabeli V. Mimo, że spośród hipotez prostych zawartych w H_e można wybrać wówczas hipotezę H_1 o wiarygodności dowolnie bliskiej 1, to i wiarygodność hipotezy H_o jest nie do zbagatelizowania (np. $\varepsilon = 0,25$). W takim przypadku test właściwie zawodzi. Dobrze jest zbadać, jak wówczas zależy wariancja empiryczna zmiennej t od frakcji córek, które uznamy za dobre kury. Jeśli któraś z tych wariancji okaże się istotnie nadnormalna, można (z pewnym ryzykiem) przyjąć H_e . Jeśli nie — trudno sprawę rozstrzygnąć. Trudność ta wynika stąd, że prostej hipotezie H_o przeciwstawiamy hipotezę złożoną H_e , z której konkurencyjną prostą hipotezę H_1 powinno się wybrać a priori, w zależności od H_e i H_o , a nie zależnie od wyników doświadczenia, tj. od empirycznej wariancji.

8. Dalsze uwagi. W praktyce, jak to zobaczymy w ustępie 9, bardzo dużo informacji zawartych w materiałach stadek selekcyjnych tracimy, ze względu na przyjętą definicję potomstwa. Jeśli mianowicie przez

potomstwo rozumiemy k (np. $k = 5$) córek jednej kury, to tracimy wszystkie informacje o kurach mających mniej niż k córek, oraz część informacji o kurach mających więcej niż k córek. Można co prawda spośród rodzeństw liczących co najmniej $2k$ córek wybrać dwa potomstwa (ewentualnie więcej), i tak w praktyce robiłem, ale wykorzystuje się w ten sposób nie wszystkie informacje, a poza tym popelnia się pewną niewłaściwość statystyczną, mianowicie losowanie kilku potomstw z jednego rodzeństwa nie jest losowaniem niezależnym. Jest ono poprawne w przypadku sprawdzania hipotezy o normalności wariancji, ale zmniejsza nieco moc testu statystycznego do porównywania wariancji empirycznej z teoretyczną.

Próbowałem (z pomocą S. Zubrzyckiego) skonstruować odpowiedni test probabilistyczny dla potomstw bez ograniczenia ilości córek w potomstwie. Okazało się jednak, że znacznie większa jest wariancja odpowiedniej zmiennej losowej t , równej liczbie dobrych córek w dowolnie liczonym potomstwie i to zarówno dla prawdopodobieństwa dobrej córki równego 0,5, jak i dla zmieniających się prawdopodobieństw. Powiększenie wariancji wskutek epistazy nie jest tak wyraźne, jak w przypadku potomstwa o stałej liczebności k . Test okazał się zbyt słaby do praktycznego wykorzystania. Uważam jednak, że celowe są poszukiwania lepszego testu do rozróżnienia między epistazą a over-dominacją, testu wykorzystującego całą informację zawartą w obserwacjach.

Ciekawe wyniki można uzyskać zmieniając liczebność potomstwa k od $k = 1$ do $k = 6$. W następnym ustępie przedstawię wyniki takiej analizy jednego z liczbowych przykładów. Nie badałem, czy wyniki te można przewidzieć teoretycznie, ani czy dla innych przykładów są one podobne; nie należy to bowiem bezpośrednio do badanego przeze mnie zagadnienia.

Przykłady

W doświadczalnych fermach drobiu wybiera się każdego roku najlepsze spośród dwuletnich kur (o sprawdzonej, dużej nieśności) i tworzy się z nich tzw. stadka selekcyjne, złożone z koguta i kilku do kilkunastu kur. Kury należące do tych stadek selekcyjnych uznajemy za dobre. Kontroluje się w ciągu roku nieśność każdej córki każdej kury-matki należącej do stadka selekcyjnego.

Dzięki uprzejmości prof. dr H. Bączkowskiej oraz prof. dr J. Kotlińskiego otrzymałem pełne materiały dotyczące stadek selekcyjnych fermy drobiu Instytutu Zootechnicznego w Czechnicy z roku 1954 i z roku 1958 oraz w Chorzelowie z roku 1958. Z materiałów tych wybrałem potrzebne mi dane.

9. Czechnica 1954. Tablica IX przedstawia te dane dla stadek selekcyjnych z Czechnicy w 1954 roku.

Liczby podkreślone w tabeli oznaczają nieśności córek wybranych do potomstw (po 5 córek). Jak zaznaczyłem wyżej, wybrano po 2 potomstwa kur Nr 508 i 842. Wszystkich podkreślonych liczb jest w tabeli 60 na ogólną liczbę 137 córek. A więc okazuje się, że przy badaniu potomstw o liczebności $k = 5$ córek, tracimy więcej niż połowę informacji.

W tabelicach X i XI podane są szeregi rozdzielcze nieśności matek i córek.

TABLICA X

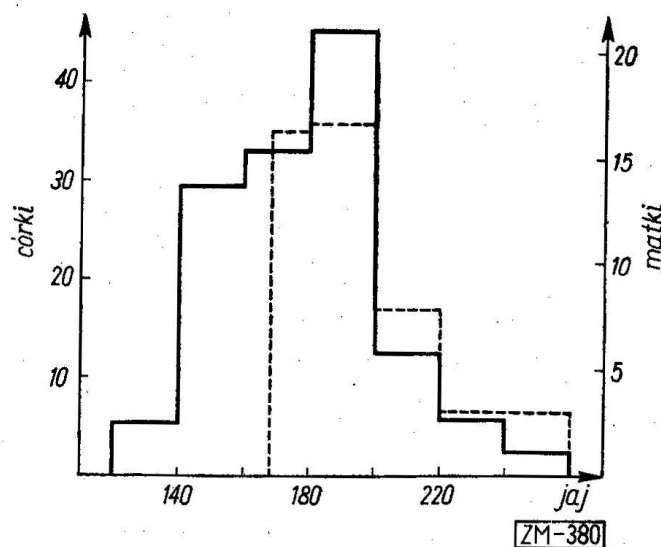
Nieśności matek

Roczna nieśność	Liczba kur
168–180	10
181–200	17
201–220	8
221–240	3
ponad 240	3
Razem	41

TABLICA XI

Nieśności córek

Roczna nieśność	Liczba kur
do 140	6
141–160	30
161–180	34
181–200	45
201–220	13
221–240	6
ponad 240	3
Razem	137



Rys. 1

Rysunek 1 przedstawia histogramy nieśności matek (linią przerywaną, i nieśności córek (linią ciągłą). Dla łatwiejszego porównania graficznego) pionową podziałkę dla histogramu matek odpowiednio powiększono.

Z porównania szeregów rozdzielczych i histogramów dla matek i córek wynika, że matki zostały odpowiednio wyselekcjonowane (o nieśności co najmniej 168 jaj rocznie). Medianą nieśności córek jest 180, a średnią arytmetyczną 178,6. Przyjmując, że wskutek selekcji nieśność córek była lepsza niż nieśność matek, można uważać, że na matki zostały wybrane

spośród wszystkich sióstr te, których nieśność przekraczała medianę. Można więc przyjąć, że wszystkie matki są, w myśl naszej definicji, dobrymi kurami.

W tablicy XII przepisałem z tablicy IX potomstwa po $k = 5$ córek oraz liczby t_i dla różnych frakcji córek uznanych za dobre. Należy zauważyć, że wśród nieśności podanych w tej tablicy medianą jest nie 180 (jak wśród wszystkich córek z tablicy IX) lecz 182.

TABLICA XII

Potomstwa kur-matek ze stadek selekcyjnych w Czechnicy w r. 1954

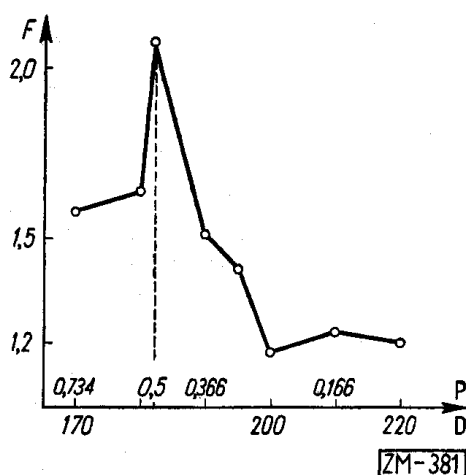
Liczba dobrych kur w potomstwie, jeśli przez dobre oznaczono kury o nieśności większej od D , stanowiące frakcję p

Nr	Roczne nieśności												
						$D = 170$	180	182	190	195	200	210	220
						$p = 0,734$	0,534	0,500	0,366	0,316	0,234	0,166	0,082
1	145	160	168	171	180	2	0	0	0	0	0	0	0
2	146	151	169	179	182	2	1	0	0	0	0	0	0
3	138	158	177	178	186	3	1	1	0	0	0	0	0
4	147	158	197	205	210	3	3	3	3	3	2	0	0
5	161	180	183	199	204	4	3	3	2	2	1	0	0
6	171	173	192	195	200	5	3	3	3	1	0	0	0
7	176	188	192	230	233	5	4	4	3	2	2	2	2
8	179	190	197	217	242	5	4	4	3	3	2	2	1
9	88	154	165	182	216	2	2	1	1	1	1	1	0
10	171	175	203	214	245	5	3	3	3	3	3	2	1
11	183	186	187	187	254	5	5	5	1	1	1	1	1
12	157	162	199	217	220	3	3	3	3	3	2	2	0
Średnia wartość \bar{t}						3,67	2,67	2,50	1,83	1,58	1,17	0,83	0,42
Wariancja empiryczna s_t^2													
(obciążona)						1,40	1,89	2,42	1,64	1,41	0,97	0,81	0,41
Wariancja nieobciążona s_t^2						1,53	2,06	2,64	1,79	1,54	1,06	0,88	0,45
Wariancja teoretyczna $5pq$						0,98	1,25	1,25	1,17	1,09	0,89	0,71	0,37
Stosunek wariancji empirycznej do teoretycznej F						1,57	1,65	2,11	1,53	1,41	1,19	1,24	1,21

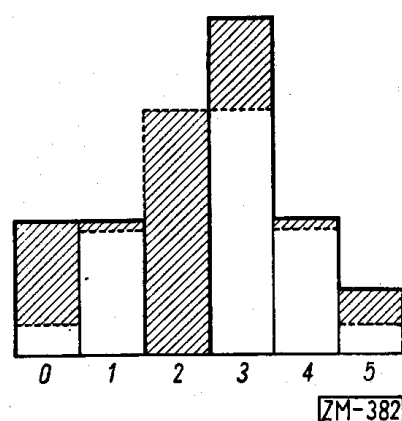
Wyżej już stwierdziliśmy, że wariancja liczby dobrych kur w potomstwie jest istotnie nadnormalna, jeśli przez dobre kury rozumiemy kury o nieśności ponad $m = 182$ jaja rocznie (stanowiące 50%). Z tablicy XII wynika, że tylko dla tej frakcji dobrych kur wariancja jest istotnie nadnormalna (przy poziomie istotności 5% i $n = 12$, stosunek F powinien przekraczać 1,8). Dla wszelkich innych frakcji dobrych kur stosunek F jest mniejszy od 1,8. Wariancja we wszystkich tych przypadkach jest nadnormalna, lecz nie istotnie.

Rysunek 2 przedstawia zależność stosunku F od frakcji dobrych kur p . Wydaje się, że maksimum tego stosunku dla $p = 50\%$ nie jest rzeczą przypadkową.

Zgodnie z wynikami ustępu 7 stwierdzamy, że na fermie w Czechnicy w 1954 roku dobre kury zawdzięczają swą wysoką nieśność epistazie. Rysunek 3 przedstawia empiryczny histogram rozkładu liczb t dobrych



Rys. 2



Rys. 3

kur w potomstwie (linią ciągłą). Dane do tego rysunku zaczerpnięte są z tablicy II. Linia przerywaną narysowany jest na tym rysunku histogram teoretyczny, tj. histogram rozkładu binomialnego dla $p = 0$ i $k = 5$. Różnice między obu rozkładami są zakreskowane. Jak widać, w 4 bocznych klasach (po 2 z każdej strony) liczebności empiryczne górują nad oczekiwanymi. Potwierdza to hipotezę o epistazie, tj. o zmianie parametru p od potomstwa do potomstwa.

10. Czechnica 1958. Tablica XIII, analogicznie do tablicy XII, przedstawia potomstwa wybrane z materiałów tejże fermi w Czechnicy z r. 1958.

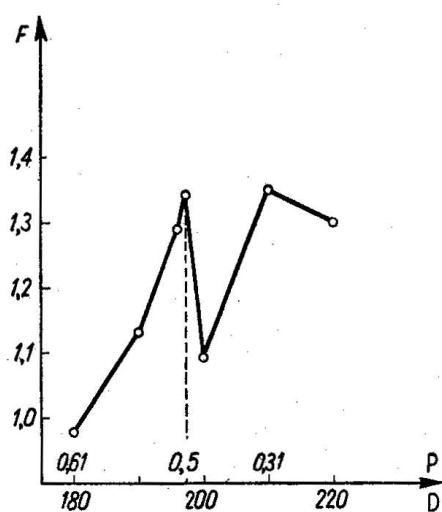
Na to, żeby stwierdzić istotną nadnormalność empirycznej wariancji trzeba by, żeby przewyższała ona wariancję teoretyczną co najmniej 1,6 razy (tyle bowiem wynosi krytyczny stosunek F Snedecora dla $n = 20$ i istotności $\varepsilon = 0,05$). Jak widać, żadna z liczb ostatniego wiersza nie przekracza tej wartości (patrz rys. 4), a więc nie można stwierdzić żeby wariancje empiryczne były istotnie nadnormalne. Wariancja empiryczna jest w tym przypadku daleko bliższa normalnej niż o 4 lata wcześniej, tj. w r. 1954. Na rysunku 5 rozkład liczby t dobrych kur w potomstwie ($k = 5$) porównany jest z rozkładem binomialnym. Jak widać, różnią się one znacznie.

Jak można zinterpretować tę pozornie dziwną niezgodność wyników w tej fermie w czteroletnim odstępie czasu? Otóż można ją wyjaśnić na-

TABLICA XIII

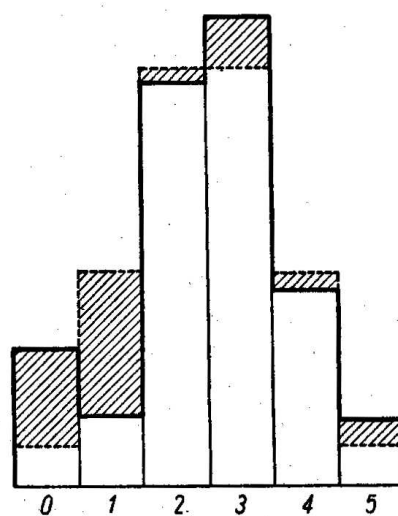
Potomstwa kur-matek ze stadek selekcyjnych w Czechnicy w roku 1958

Nr	Roczne nieśności					Liczba dobrych kur-córek							
						$D =$	180	190	195	196	200	210	220
						$p =$	0,610	0,580	0,520	0,510	0,450	0,310	0,190
1	133	151	192	201	209	3	3	2	2	2	0	0	
2	156	213	227	232	246	4	4	4	4	4	4	3	
3	162	199	208	218	231	4	4	4	4	3	2	1	
4	135	170	176	203	222	2	2	2	2	2	1	1	
5	178	179	188	209	221	3	2	2	2	2	1	1	
6	177	183	200	204	213	4	3	3	3	2	1	0	
7	153	154	227	229	234	3	3	3	3	3	3	3	
8	126	135	169	174	188	1	0	0	0	0	0	0	
9	145	175	204	206	220	3	3	3	3	3	1	0	
10	176	191	194	226	239	4	4	2	2	2	2	2	
11	158	162	196	208	220	3	3	3	2	2	1	0	
12	171	174	175	191	225	2	2	1	1	1	1	1	
13	138	159	180	198	208	2	2	2	2	1	0	0	
14	137	144	155	180	192	1	1	0	0	0	0	0	
15	144	167	212	222	230	3	3	3	3	3	3	2	
16	198	202	203	208	233	5	5	5	5	4	1	1	
17	180	194	197	214	217	4	4	3	3	2	2	0	
18	149	205	211	232	240	4	4	4	4	4	3	2	
19	131	189	213	218	231	4	3	3	3	3	3	1	
20	133	160	197	214	229	3	3	3	3	2	2	1	
Średnia wartość t						3,10	2,90	2,60	2,55	2,25	1,55	0,95	
Wariancja obciążona s_t^2						1,09	1,59	1,54	1,55	1,29	1,35	0,95	
Wariancja nieobciążona s_t^2						1,15	1,67	1,62	1,63	1,36	1,42	1,00	
Wariancja teoretyczna $5pq$						1,18	1,22	1,25	1,25	1,24	1,07	0,77	
Stosunek wariancji empirycznej do teoretycznej F						0,97	1,12	1,29	1,31	1,09	1,33	1,29	



Rys. 4

ZM-383



Rys. 5

ZM-384

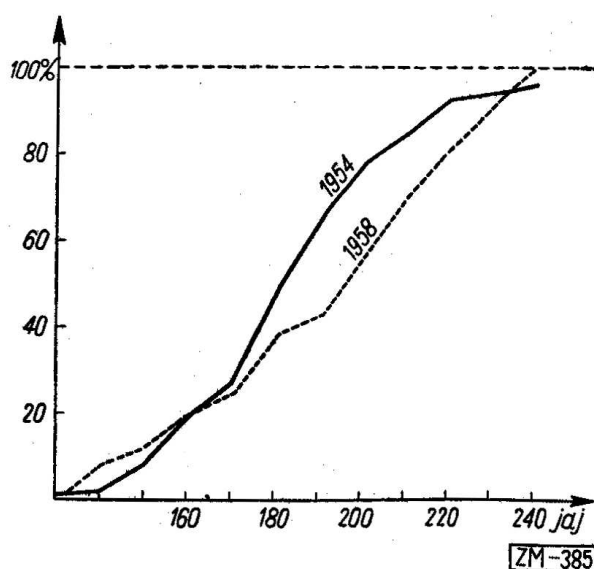
stepująco: W roku 1954 mieliśmy w Czechnicy do czynienia z epistazą. Dobre kury charakteryzowały się zespołem genów AABB (wyłącznie, lub między innymi). W sytuacji epistatycznej można, jak wiadomo, drogą selekcji podnieść dobroć kur, tj. wybrać z całej populacji kury o zespole genów AABB na wyróżnionych lokusach. Taki wyselekcjonowany skład populacji jest dziedzicznie stabilny. Otóż właśnie taka selekcja miała miejsce w Czechnicy. Widać to z porównania z tablic XII i XIII. W tablicy XII na 60 nieśności jest tylko 14 przekraczających 200 jaj rocznie, czyli 23,3%, w tablicy XIII na 100 nieśności jest 45 przekraczających 200 jaj rocznie. Podobnie obliczyłem inne pozycje zawarte w tablicy XIV.

TABLICA XIV

Fracje (w procentach) kur-córek o nieśności nie przekraczającej D jaj rocznie na fermie w Czechnicy w roku 1954 i w roku 1958

$\begin{array}{c} D \\ \text{Rok} \end{array}$	130	140	150	160	170	180	190	200	210	220	230	240
1954	1,7	3,3	8,3	20,0	26,7	46,7	63,3	76,7	83,3	91,7	93,3	95,0
1958	1,0	8,0	12,0	20,0	25,0	38,0	42,0	55,0	69,0	81,0	91,0	99,0

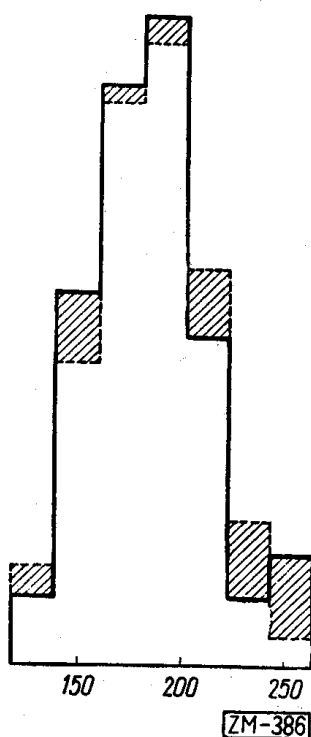
Rysunek 6 przedstawia wykresy tej tablicy, czyli ogiwy (dystrybuanty) nieśności kur-córek w latach 1954 (ciągłą linią) i 1958 (przerywaną).



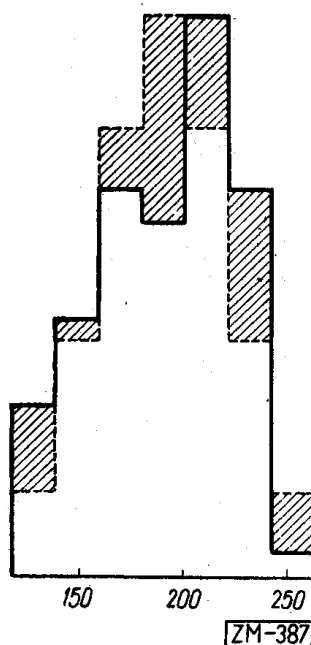
Rys. 6

Ogiwa nieśności dla roku 1954 ma w przybliżeniu kształt dystrybuanty rozkładu normalnego. Natomiast ogiwa dla roku 1958 ma kształt bardzo charakterystyczny, niemal liniowy, odbiegający znacznie od nor-

malnego. Istotne różnice występują szczególnie w nieśnościach od 180 do 230 jaj rocznie. Takich kur jest w roku 1958 istotnie więcej, niż w r. 1954. Natomiast dla nieśności poniżej 180 oraz powyżej 230, ogiwy obu lat nie są zbyt różne. Różnica między dwiema ogiwami na rysunku 6 jest wynikiem czteroletniej selekcji. Z populacji w przybliżeniu normalnej (w 1954) selekcja ta uczyniła populację odbiegającą od normalnej, w której dopatrzeć się można niejednorodności, rozbitcia na części. Można przypuszczać, że jedną z tych części tworzą kury o pożądanym zespole genów na wyróżnionych lokusach, kury o wysokiej nieśności, a mianowicie o nieśności powyżej 180 jaj rocznie. Ilość tych właśnie kur wzrosła w ciągu czterech lat selekcji.



Rys. 7



Rys. 8

Aby porównać dokładniej nieśności kur fermy w Czechniey w roku 1954 i 1958, porównałem na rysunkach 7 i 8 rozkłady empiryczne nieśności z rozkładem normalnym o tej samej średniej i wariancji co rozkład empiryczny. Jak widać, w r. 1954 rozkład empiryczny niemal nie różni się od normalnego (χ^2 wynosi 2,5 przy 4 stopniach swobody). Natomiast w r. 1958 empiryczny rozkład nieśności różni się znacznie od normalnego (χ^2 wynosi 14,7 przy 6 stopniach swobody; wiarogodność hipotezy zerowej, o nieistotnej różnicy między rozkładem empirycznym a normalnym, jest mniejsza niż 5%). Zaznacza się szerokie rozciągnięcie rozkładu, a nawet drugi wierzchołek. Wskazuje to na efekt selekcji. Z jednorodnego

materiału w r. 1954 otrzymano w ciągu 4 lat selekcji materiał niejednorodny, z którego można by wydzielić część o pożądanym genotypie, czyli o większej nieśności oraz drugą część o innych genotypach, pochodzących z rozszczepiania się mieszańców genotypowo dobrych wskutek dominacji, lub z kogutów o nie dobrych genotypach.

Można zaryzykować prognozę, że w Czechnicy selekcja, polegająca na usuwaniu kur o niskiej nieśności (poniżej 180 jaj rocznie), będzie jeszcze przez pewien czas dawała pozytywne wyniki, gdyż, jak to widzieliśmy, wariancja zmiennej t jest jeszcze trochę nadnormalna, a na rysunku 6 jest jeszcze sporo kur o nie pożądanym genotypach, tj. o nieśności poniżej 180. Po pewnym czasie selekcja doprowadzi do tego, że nie będzie na fermie w Czechnicy kur o tak niskiej nieśności, a wówczas wariancja zmiennej t (liczby dobrych kur w potomstwie) będzie normalna. Wówczas dalsza selekcja nie da już rezultatu. Będziemy mieli do czynienia z sytuacją over-dominacji i tylko metodą krzyżówki produkcyjnej będziemy mogli podnieść nieśność kur.

Należy przy tym zauważyć, że potomstwa złożone wyłącznie z dobrych kur mają z większym prawdopodobieństwem pożądaną zespół genów AABB niż dobre kury z innych potomstw. Tak np. w Czechnicy w r. 1958 specjalnie nadają się do dalszej hodowli potomstwa Nr 16 i 17. Należałoby się zastanowić, czy nie byłoby właściwe wykorzystanie jakiejś ze specjalnych metod wyhodowania czystej linii o zespole genów AABB, np. metody linii wsobnych.

11. Chorzeliów. Tablica XV, analogicznie do tablic XII i XIII, przedstawia potomstwa wybrane (z wszystkich rodzeństw liczących co najmniej 5 córek) z materiałów Fermi Zarodowej w Chorzeliowie z roku 1958.

Tym razem, jak widać, wariancja empiryczna nie różni się istotnie od teoretycznej. Stosunek F nieznacznie różni się od jedności z wyjątkiem jednej tylko frakcji dobrych kur-córek $p = 0,636$. W tym zresztą przypadku $F = 1,21$, czyli również wskazuje na nieistotną różnicę między wariancją empiryczną a teoretyczną. W szczególności, dla $p = 0,51$ (nie można było uzyskać $p = 0,5$) $F = 0,96$, czyli prawie 1. Nasz test niemal idealnie wskazuje na over-dominację. Z założenia over-dominacji wynika nie tylko to, że empiryczna wariancja powinna być normalna, lecz także, że rozkład prawdopodobieństw zmiennej losowej t powinien być rozkładem binomialnym. W tablicy XVI porównane są dwa rozkłady: empiryczny rozkład frekwencji poszczególnych wartości zmiennej t (wzięty z tablicy XV) i teoretyczny rozkład prawdopodobieństw, tj. binomialny z parametrami $p = 0,5$ i $k = 5$. Dla łatwiejszego porównywania prawdopodobieństwa te zostały pomnożone przez 33 (są to więc oczekiwane frekwencje).

TABLICA XV

Potomstwa kur-matek ze stadek selekcyjnych w Chorzelowie

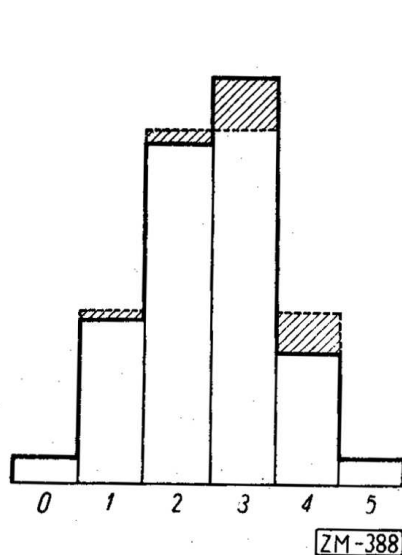
Nr	Roczna nieśność						Liczba dobrych kur-córek							
							$D = 130$	140	149	160	170	175	180	190
							$p = 0,752$	0,636	0,510	0,350	0,210	0,190	0,140	0,100
1	99	112	146	148	155	3	3	1	0	0	0	0	0	0
2	146	160	168	179	192	5	5	4	3	2	2	1	1	1
3	117	122	127	154	179	2	2	2	1	1	1	0	0	0
4	100	112	169	177	197	3	3	3	3	2	2	1	1	1
5	107	130	147	162	168	3	3	2	2	0	0	0	0	0
6	114	151	170	183	200	4	4	4	3	2	2	2	1	1
7	110	135	138	149	152	4	2	1	0	0	0	0	0	0
8	141	148	151	153	175	5	5	3	1	1	0	0	0	0
9	111	127	146	187	225	3	3	2	2	2	2	2	1	1
10	129	141	160	162	175	4	4	3	2	1	0	0	0	0
11	141	146	150	163	211	5	5	3	2	1	1	1	1	1
12	115	135	142	160	166	4	3	2	1	0	0	0	0	0
13	117	155	161	168	251	4	4	4	3	1	1	1	1	1
14	110	128	156	162	186	3	3	3	2	1	1	1	0	0
15	114	120	123	128	135	1	0	0	0	0	0	0	0	0
16	132	140	152	168	177	5	3	3	2	1	1	0	0	0
17	131	136	157	157	162	5	3	3	1	0	0	0	0	0
18	127	141	143	193	216	4	4	2	2	2	2	2	2	2
19	113	114	134	148	180	3	2	1	1	1	1	0	0	0
20	123	127	135	138	170	3	1	1	1	0	0	0	0	0
21	106	138	142	162	186	4	3	2	2	1	1	1	0	0
22	113	129	134	166	209	3	2	2	2	1	1	1	1	1
23	143	153	170	186	200	5	5	4	3	2	2	2	1	1
24	160	169	177	200	206	5	5	5	4	3	3	2	2	2
25	125	136	166	177	210	4	3	3	3	2	2	1	1	1
26	139	147	156	161	196	5	4	3	2	1	1	1	1	1
27	133	140	146	149	178	5	3	1	1	1	1	0	0	0
28	111	128	132	153	189	3	2	2	1	1	1	1	0	0
29	120	142	155	156	158	4	4	3	0	0	0	0	0	0
30	118	130	143	195	196	3	3	2	2	2	2	2	2	2
31	112	127	135	157	159	3	2	2	0	0	0	0	0	0
32	116	146	166	175	183	4	4	3	3	2	1	1	0	0
33	95	122	152	161	164	3	3	3	2	0	0	0	0	0
Średnia wartość \bar{t}							3,76	3,18	2,55	1,73	1,03	0,94	0,70	0,48
Wariancja obciążona s_t^2							0,97	1,36	1,16	1,11	0,70	0,72	0,57	0,43
Wariancja nieobciążona s_t^2							1,00	1,40	1,20	1,14	0,72	0,74	0,59	0,44
Wariancja teoretyczna $5pq$							0,93	1,16	1,25	1,14	0,83	0,77	0,60	0,45
Stosunek wariancji empirycznej do teoretycznej F							1,08	1,21	0,96	1,00	0,87	0,97	0,98	0,99

TABLICA XVI

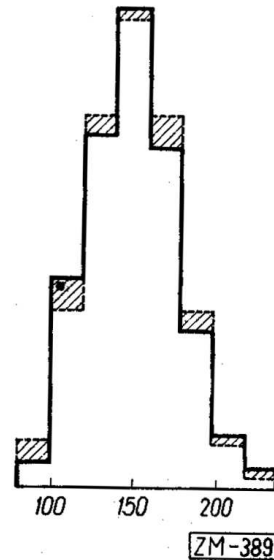
Zaobserwowane i oczekiwane frekwencje poszczególnych wartości zmiennej losowej t

Wartości zmiennej t	0	1	2	3	4	5
Zaobserwowane frekwencje	1	5	10	12	4	1
Teoretyczne frekwencje	1,0	5,2	10,3	10,3	5,2	1,0

I tutaj zgodność między rozkładem empirycznym a teoretycznym, przewidywanym a priori w przypadku over-dominacji, jest niemal idealna. Widać to z rysunku 9 w porównaniu z analogicznymi rysunkami 3 i 5.



Rys. 9



Rys. 10

Rysunek 10, analogiczny do rysunków 7 i 8, przedstawia histogram nieśności kur-córek z fermi w Chorzelowie (linią ciągłą) oraz odpowiadający mu histogram rozkładu normalnego (linią przerywaną). Różnice między obu histogramami są zakreskowane. Jak widać, rozkład nieśności tych kur jest bardzo podobny do normalnego, bardziej niż to miało miejsce dla kur z fermi Czechnickiej w r. 1954, nie mówiąc już o kurach selekcyjowanych przez 4 lata. Z porównania 3 empirycznych histogramów z normalnymi na rysunkach 7, 8 i 10 można wnosić, że w populacji kur, w której nie prowadzono właściwej, tj. udanej, selekcji podnoszącej istotnie nieśność, rozkład nieśności jest w przybliżeniu normalny. Przy czym w sytuacji over-dominacji przybliżenie to jest lepsze, a w sytuacji epistazy — gorsze. Natomiast w populacji poddanej właściwej selekcji (ale nie zakończonej) rozkład nieśności znacznie się różni od normalnego.

Wobec takiej sytuacji, na fermie kurzej w Chorzelowie nie należy się spodziewać dobrych wyników zwykłej selekcji. Nie może ona doprowadzić do genetycznie stabilnego podniesienia nieśności, gdyż z powodu over-dominacji najlepszą nieśność dają niestabilne heterozygoty. Można by tam wyselekcjonować dwie czyste linie o obu homozygotycznych zespołach genów i osiągnąć wyższą nieśność ich krzyżówki (produkcyjnej), jednakże i ta nieśność nie będzie zbyt wysoka, bo kury o heterozygotycznym układzie genów, tj. dobre, stanowią około 50% kur-córek w potomstwach

(patrz tablica XV), a średnia nieśność tych dobrych kur wynosi 173,5 jaja rocznie.

12. Próby dla potomstw o innej liczebności k . Można sobie zadać pytanie, czy wyniki przedstawione w przykładach niniejszej pracy nie są dziełem przypadku. Jak wiadomo, potomstwa zapisane na tablicach XII, XIII i XV zostały wybrane z liczebniejszych potomstw. Zastanawialiśmy się już nad tym, czy nie można by wykorzystać wszystkich danych zawartych w obserwacjach, tj. wszystkich potomstw w takich liczebnościach jakie naprawdę występują. Okazało się jednak, że teoretyczny rozkład liczby dobrych kur-córek w takim potomstwie o zmiennej liczebności, jest bardziej rozciągnięty niż rozkład binomialny (ma wariancję nadnormalną) i dlatego kryterium mające odróżniać rozkład empiryczny od teoretycznego jest daleko słabsze.

Można jednak zastanawiać się nad tym, czy liczebność potomstwa $k = 5$ jest najwłaściwsza, czy nie należałoby budować testu na większych lub mniejszych liczebnościach. Im większa będzie liczebność potomstwa k , tym lepsza powinna być estymacja wariancji empirycznej w stosunku do teoretycznej, czyli tym mniejszy błąd II rodzaju. Wynika stąd, że powinniśmy dążyć do jak największych liczebności k . Z drugiej jednak strony im większe k , tym mniej potomstw o liczebności k można wybrać z materiału, czyli tym więcej tracimy informacji. Należy więc zbadać jak zmniejsza się ilość potomstw ze wzrostem liczebności k i na tej podstawie ustalić najwłaściwszą liczebność.

W tablicy IX podane są wszystkie potomstwa kur należących do stadek selekcyjnych w fermie Czechnickiej w roku 1954. Poniżej, w tablicy XVII, przedstawiam liczbę potomstw n o liczebności k , jakie można wybrać z tablicy IX.

TABLICA XVII

Liczba n potomstw o liczebności k oraz wariancje liczby dobrych kur w tych potomstwach, wybranych spośród kur-matek należących do stadek selekcyjnych w Czechnicy w r. 1954

k	1	2	3	4	5	6	7	8	9	10	11
n	137	54	31	19	12	5	4	3	3	2	0
me	180	179	182	179	182	188	191	193	192	191	
Et	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	
\bar{t}	0,5	1,0	1,52	2,0	2,5	3,0	3,5	3,67	4,67	5,0	
σ_t^2	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50	
s_t^2	0,25	0,57	0,89	1,33	2,64	1,50	1,67	0,34	0,22	2,00	
F	1,00	1,14	1,19	1,33	2,11	1,00	0,95	0,17	0,10	0,80	
F_{kryt}	1,21	1,32	1,46	1,60	1,79	2,37	2,60	3,00	3,00	3,84	

Dla $k = 1$ wariancja empiryczna musi być normalna, więc takich potomstw nie można brać pod uwagę przy odróżnianiu badanych hipotez H_o i H_e .

Dla $k = 6, 7, 8, 9$ i 10 liczba wybranych potomstw o liczebności k nie przekracza 5. Empiryczna wariancja obliczona z tak małej liczby obserwacji nie przedstawia praktycznie żadnej wartości. Np. dla $k = 9$ wariancja teoretyczna jest 10,2 razy większa od empirycznej, ale nie przeczy to hipotezie normalności, ani nawet nadnormalności wariancji empirycznej na poziomie istotności 5%, gdyż wariancja teoretyczna ma prawo być jeszcze większa w porównaniu z empiryczną (19,5 razy).

Dla liczebności $k = 2, 3, 4$ i 5 ilości potomstw n są rzędu kilkunastu do kilkudziesięciu. Krytyczne wartości F zmieniają się od 1,32 do 1,79. Jak widać ze wzrastania stosunku F (1,14 1,19 1,33 2,11) powiększanie liczebności k , mimo zmniejszania się ilości potomstw n , daje coraz wyraźniejszą różnicę między empiryczną a teoretyczną wariancją. Dalsze powiększanie liczebności k pociąga silny spadek ilości potomstw n . Tak np. przy powiększeniu k z 5 do 6, następuje spadek liczby potomstw z 12 do 5, czyli o 58%.

Wydaje się, że zjawisko to jest biologicznym. Ilość jaj jednej kury, które można przeznaczyć do inkubacji, straty w inkubatorze, oddzielenie kogutków i wreszcie straty w ciągu blisko dwuletniej hodowli (w celu zbadania nieśności rocznej) powodują, że najczęściej liczebność potrzebnego nam potomstwa nie przekracza $k = 5$.

Można jeszcze zapytać, czy duża wariancja zmiennej t dla $k = 5$ nie jest rzeczą przypadku, tj. czy otrzymalibyśmy również nadnormalną wariancję, gdybyśmy wybrali potomstwa inaczej niż to pokazane w tablicy IX. Aby i to zbadać wybrałem z tablicy IX inne 5-sztukowe potomstwa, i aby otrzymać jak najgorszy wynik, wybrałem je zależnie od potomstw podanych w tablicy IX. Mianowicie, wybrałem do nowych potomstw przede wszystkim sztuki pominięte w poprzednim wyborze i uzupełniałem je do 5 sztuk losowo. Otrzymałem $s_i^2 = 2,45$ i $F = 1,96$, więc nieco mniejsze od poprzednio otrzymanych, lecz również istotnie nadnormalne.

Serdecznie dziękuję prof. dr H. Bączkowskiej, kierowniczce Działu Hodowli Drobiu Instytutu Zootechnicznego w Krakowie oraz prof. dr J. Kotlińskiemu, dyrektorowi Zakładu Doświadczalnego tegoż Instytutu w Czechnicy za udostępnienie mi potrzebnych materiałów. Niemniej wdzięczny jestem kierowniczkom Działu Hodowli Zarodowej Drobiu mgr H. Ciechanowskiej i mgr T. Kwiatkowskiej w Czechnicy za przygotowanie tych materiałów, wiele wyjaśnień i informacji, które umożliwiły mi należyte ich wykorzystanie.

Prace cytowane

- [1] J. L. Lush, *Animal breeding plans*, Ames Collegiate Press, 1945.
- [2] J. Perkal, *Matematyka dla przyrodników i rolników II*, w druku.
- [3] A. M. Srb i R. D. Owen, *Genetyka ogólna*, PWRiL, 1959.
- [4] H. Steinhaus, *Elementary inequalities between the expected values of current estimates of variance*, Coll. Math. 1 (1948), str. 312-321.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

Praca wpłynęła 21. 2. 1961

Ю. ПЕРКАЛЬ (Вроцлав)

ОБ ОДНОМ ГЕНЕТИЧЕСКОМ КРИТЕРИИ

РЕЗЮМЕ

Сверхдоминантность это взаимодействие между двумя генами одного локуса, при котором гетерозыготный набор генов Аа даёт более высокий генотиповый эффект, чем гомозыготные наборы АА и аа. Эпистаз заключается во взаимодействии генов в двух локусах, при котором набор генов, например ААВВ, даёт более высокий генотиповый эффект, чем любые другие наборы генов в тех же локусах. Для поднятия генетического эффекта в случае эпистаза следует применять обыкновенный отбор, который приведёт к образованию генетически стабильной породы, обладающей более высоким генетическим эффектом. В случае сверхдоминантности такой отбор невозможен. Животновод должен вырастить две чистые линии, скрещивание которых в первом поколении приведёт к явлению сверхдоминантности. В связи с изложенным большое практическое значение имеет критерий, позволяющий отличить сверхдоминантность от эпистаза. В настоящей работе предлагается первый критерий такого рода. В изложении используется пример куриной фермы.

Критерий. Курицу называем хорошей, если количество яиц, которые она сносит в течение года, больше средней (медианы). Обозначаем через n количество хороших куриц, имеющих не менее пяти дочерей, яйценоскость которых была проверена. Эти пять дочерей хорошей курицы и произвольного петуха будем называть потомством (если курица имела более пяти дочерей, производим случайную выборку потомства). Символом t_i ($i = 1, 2, \dots, n$) обозначаем количество хороших куриц — дочерей среди потомства i -ой курицы. Если эмпирическая дисперсия чисел t_i не отличается существенно от числа 1,25, то имеем дело со сверхдоминантностью. Если эмпирическая дисперсия чисел t_i существенно больше 1,25, то в исследуемой ферме имеем дело с эпистазом.

Таблицы II, III и IV изображают распределения чисел t_i в трёх случаях: Чехница 1954, Чехница 1958 и Хожелёв. В первом случае $s_t^2 = 2,64$, следовательно существенно больше, чем 1,25: имеем здесь явление эпистаза. Во втором случае $s_t^2 = 1,63$: положение неопределённое. В третьем $s_t^2 = 1,20$, т. е. четкое явление сверхдоминантности.

Обоснование критерия. В случае сверхдоминантности предполагаем, что хорошими являются те курицы, которые имеют в выделенном локусе набор генов Аа. Таблица V показывает, каким образом наследуется набор генов по хорошей матери и произвольном петухе. Оказывается (см. последнюю строку этой таблицы), что вероятность того, чтобы дочь хорошей курицы была также хорошей курицей равна 0,5 не зависимо от петуха. Следовательно, случайная величина t , т. е. количество хороших дочерей в потомстве, насчитывающим 5 дочерей хорошей матери и произвольного петуха, имеет биномиальное распределение с вероятностью $p = 0,5$ и длиной серии $k = 5$. Математическое ожидание случайной величины t равно $kp = 2,5$, а дисперсия этой величины $kp(1-p) = 1,25$. Следовательно, $p = 50\%$ всех куриц одного поколения это хорошие курицы (ясно те, которые имеют большую яйценоскость), а опытная дисперсия чисел t_i не должна существенно отличаться от 1,25.

В случае эпистаза обозначая цифрами следующие наборы генов в двух локусах:

$$1 = AABV, 2 = AABb, 3 = AAbb, 4 = AaBV, 5 = AaBb, 6 = Aabb, 7 = aaBV, \\ 8 = aaBb, 9 = aabb.$$

Хорошими будем в этом случае считать курицы со следующими наборами генов: а. 1 или b. 1 и 2 (т. е., когда удостоверимся, что ген В доминирует над геном b или с. 1 и 4 (когда ген А доминирует над геном а или d. 1, 2, 4, 5 (когда ген А доминирует над а и ген В над b). В таблице VII показано, каким образом наследуются наборы генов в выделенных двух локусах по хорошим матерям (в смысле а, b, с или d) и произвольным петухам. Таблица VIII содержит вероятности того, что дочь хорошей матери будет хорошей курицей (для каждого из случаев а, b, с, d — особая таблица). Мы видим, что эти вероятности изменяются от 0 до 1 в зависимости от петуха и курицы (в случаях b, с и d). Следовательно, доля хороших куриц среди всех куриц какого-нибудь поколения не обязательно равна 50%, а упомянутая выше случайная величина t , равная количеству хороших куриц в потомстве насчитывающем 5 дочерей, не имеет биномиального распределения. Распределение этой случайной величины соответствует схеме Лексиса, и, следовательно, её дисперсия будет сверхнормальная.

Сформулированный в начале критерий следует из сравнения достоверности гипотез о сверхдоминантности и эпистазе.

Примеры. Таблица IX представляет яйценоскость матерей (в третьем столбце) и дочерей (в следующих столбцах) куриц с фермы в Чехницы в 1954 г. На черт. 1 сравниваются распределение носкости матерей (пунктир) избранных в племенные стада с распределением носкости дочерей (сплошная линия). Учитывая эффект отбора, можем считать, что матери являются хорошими курицами своего поколения. В таблице XII приведены носкости 60 дочерей (подчёркнутые в таблице IX), являющихся случайно выбранными 12 потомствами по 5 дочерей. В шестом от конца столбце приведены значения случайной величины t , т. е. числа хороших дочерей в отдельных потомствах, при условии, что хорошими будем считать $p = 0,50$ дочерей с лучшей носкостью. В третьей снизу строке приведена дисперсия величины t равная 2,64. В последней строке указано её отношение $F = 2,11$ к теоретической дисперсии. Это отношение свидетельствует о том, что опытная дисперсия этой величины действительно сверхнормальная. На черт. 3 сравнивается распределение величины t (сплошная линия) с биномиальным распределением (пунктир). Заштрихованные расхождения значительны. Другие столбцы правой части таблицы XII содержат значения величины t и соответствующ-

щие показатели в случае, если хорошими дочерями считать не 0,5 но $p = 0,734, 0,534, \dots, 0,082$ куриц с лучшей яйценоскостью. Черт. 2 изображает зависимость отношения F от p . Мы видим, что F везде больше единицы (т. е. для каждого p опытная дисперсия величины t сверхнормальна), причём F принимает наибольшее значение при $p = 0,5$.

Таблица XIII аналогична предыдущей. Она изображает 20 потомств выбранных с той же фермы в Чехницы 4 года спустя, т. е. в 1958 г. Здесь также дисперсия сверхнормальна для всех p (за исключением $p = 0,61$), но не так существенно (см. черт. 4). Распределение величины t гораздо больше походит на биномиальное распределение (для $p = 0,5$ см. черт. 5). Причиной различия между состояниями в 1954 и 1958 годах является отбор, который проводился всё это время. Черт. 6 изображает статистические совокупности носкости дочерей в 1954 и 1958 годах. Мы видим, что в 1958 г. на ферме было больше куриц с носкостью от 180 до 230 яиц в год. На рефме выступало явление эпистаза, поэтому отбор давал хорошие результаты. В 1958 г. ещё наблюдается явление эпистаза, поэтому можно предполагать, что ещё некоторое время отбор будет давать хорошие результаты. После отбор исключит курицы с нежелательными генотипами, останутся только хорошие курицы (несущие свыше 180 яиц в год) и явление эпистаза исчезнет.

Таблица XV аналогична таблицам XII и XIII. Она изображает итоги обследования 33 потомств фермы в Хожелёве. Здесь дисперсия величины t нормальна для всех значений p (см. последнюю строку таблицы XV), для $p = 0,5$ распределение этой величины почти точно биномиальное (см. черт. 9). Следовательно, имеем дело со сверхдоминантностью. Обыкновенный отбор не может здесь привести к положительным результатам. Здесь можно только отобрать две чистые линии и производить гетерозыготную помесь, но даже при таком методе не следует рассчитывать на высокую носкость помеси (в среднем 173,5 яиц в год).

На черт. 7, 8 и 10 сравниваются распределения носкости куриц этих трёх примеров с нормальным распределением (обозначенным пунктиром). Мы видим, что наиболее походит на нормальное распределение носкости куриц в Хожелёве, наименее — куриц селекционированных в течение 4 лет в Чехницы.

Критерий, базирующийся на потомствах, насчитывающих по 5 дочерей, не использует значительного количества сведений, содержащихся в опытных данных (см. например не подчеркнутые носкости дочерей в таблице IX). Поиски критерия, который дал бы возможность учитывать все сведения т. е. сведения о потомствах различной численности, не имели успеха. Оказалось, что распределения соответствующих величин t были слишком растянуты, а их дисперсии во всех случаях сверхнормальны. Критерий для сравнения опытных дисперсий с теоретическими был слишком слаб. Таблица XVII изображает анализ различных численностей k потомства. Чем меньше k , тем хуже оценка дисперсии величины t . Чем больше k , тем меньше количество n соответствующих потомств. Поэтому кажется, что $k = 5$ является оптимальным значением в рассматриваемом случае.

J. PERKAL (Wrocław)

A DESIGN FOR A GENETIC TEST

SUMMARY

Overdomination is an interaction between two genes of one locus in which the heterozygotic combination of genes Aa has a higher genotypic value than the homozygotic combinations AA and aa . Epistasis consists in an interaction of genes in two

loci in which one combination of genes, say AABB, has a higher genotypic value than each of the other combinations of genes in those loci. In the case of epistasis the breeder, in order to increase the genotypic value, should apply ordinary selection, which will result in a genetically stable population with a high genotypic value. In the case of overdomination such selection is impossible. The breeder should breed two pure lines, whose cross-breed in the first generation will produce overdomination.

A test for distinguishing overdomination from epistasis is thus both theoretically and practically important. The present paper describes the first test of this kind, a poultry farm being used as an example.

The test. Hens which annually lay more eggs than the median are termed good. I denote by n the number of good hens having each at least five daughters whose laying capacity has been checked. Those five daughters of a good hen and any cock are termed a *progeny* (if the hen has had more than five daughters, the progeny is chosen at random).

The letter t_i ($i = 1, 2, \dots, n$) denotes the number of good daughter-hens in the progeny of the i -th hen. If the empirical variance of numbers t_i does not differ significantly from number 1.25, we are dealing with overdomination. If, on the other hand, the empirical variance of numbers t_i is significantly greater than 1.25, we are dealing with epistasis on the farm in question.

Tables II, III and IV present the distributions of numbers t_i in three cases: Czechnica 1954, Czechnica 1958 and Chorzelów. In the first case $s_t^2 = 2.64$, which is significantly greater than 1.25: we thus have the effect of epistasis. In the second case $s_t^2 = 1.63$: the situation is not clear. In the third case $s_t^2 = 1.20$ — a clear case of overdomination.

Explanation of the test. In the case of overdomination we assume that good hens are those which have in a given locus the Aa combination of genes. Table V shows how a combination of genes from a good hen and any cock is inherited. It is shown (see the last row of the table) that the probability of a daughter of a good hen being a good hen herself is 0.5 independently of the cock. Hence the conclusion that the random variable t , i.e. the number of good daughters in a progeny totalling 5 daughters of a good hen and any cock has a binomial distribution with probability $p = 0.5$ and length of series $k = 5$. The expected value of the random variable t is $kp = 2.5$ and the variance of that variable is $kp(1-p) = 1.25$. Thus $p = 50\%$ of all the hens of one generation are good hens (those of course which have greater laying capacity) and the empirical variance of numbers t_i should not significantly differ from 1.25.

In the case of epistasis I denote by figures the following combinations of genes in two loci: 1 = AABB, 2 = AABb, 3 = AAbb, 4 = AaBB, 5 = AaBb, 6 = Aabb, 7 = aaBB, 8 = aaBb, 9 = aabb. We shall regard those hens as good which have the following combinations of genes: a. 1, or b. 1 and 2 (if we find that gene B dominates over gene b, or c. 1 and 4 (if gene A dominates over gene a), or, finally, d. 1, 2, 4, 5 (if gene A dominates over gene a and gene B dominates over gene b). Table VII shows the inheritance of gene combinations in given two loci from good mothers (in the sense of a, b, c or d) and from any cock. From the Table VIII it can be seen that those probabilities vary from 0 to 1 according to the cock and the hen (in cases b, c and d). Hence the conclusion that the fraction of good hens among all the hens of a generation need not be equal to 50%, and thus the above-mentioned random variable t equal to the number of good daughters in a progeny totalling five hens does not have a binomial distribution. The distribution of that variable corresponds to the Lexis scheme and consequently its variance is greater than normal.

The test formulated at the beginning of this paper results from a comparison of the likelihoods of two hypotheses, that of overdomination and that of epistasis.

Examples. Table IX shows the laying capacities of mothers (column 3) and daughters (further columns) on the poultry farm of Czechnica in the year 1954. Fig. 1 gives a comparison of the distribution of the laying capacity of the mothers (dotted line) chosen for selected breeding with the distribution of the laying capacity of the daughters (continuous line).

Taking into consideration the effect of selection, we can regard the mothers as good hens in their generation. Table XII shows the laying capacities of 60 daughters (underlined in Table IX), namely 12 progenies (chosen at random) of 5 daughters each. The sixth column from the end contains the values of the random variable t , i.e. the numbers of good daughters in each progeny, if we regard as good $p = 0,50$ daughters with better laying capacities. In the third row from the bottom we are given the variance of the variable $t = 2,64$ and in the last row — its ratio to the theoretical variance $F = 2,11$. This ratio shows that the empirical variance of that variable is indeed greater than normal. In Fig. 3 the distribution of variable t (continuous line) is compared with a binomial distribution (dotted line). The shaded divergences are considerable. The other columns of the right-hand side of Table XII contain the values of variable t and the corresponding indices, if we regard as good daughters not the fraction 0,50 but other fractions ($p = 0,734, 0,534, \dots, 0,082$) of hens with the best laying capacity. Fig. 2 shows the dependence of ratio F on fraction p . It can be seen to be everywhere greater than 1 (i.e. for every p the empirical variance of the variable t is greater than normal); ratio F reaches its maximum value for $p = 0,5$.

Table XIII, analogous to the preceding one, presents 20 progenies selected from the same farm in Czechnica 4 years later, i.e. in the year 1958. Here the variance is also greater than normal for all fractions p (with the exception of $p = 0,61$) but less significantly so (see Fig. 4). The distribution of variable t is much more like the binomial distribution (for $p = 0,5$ see Fig. 5). The difference between this situation and the situation four years earlier should be attributed to the selection continued throughout this period. Fig. 6 shows the data (distribution functions) of the laying capacity of daughters in 1954 and in 1958. They indicate that in 1958 there were more hens on the farm with the laying capacity from 180 to 230 eggs yearly. The effect of epistasis had appeared on the farm and therefore selection was giving good results. In 1958 the effect of epistasis is still seen, and thus it can be expected that selection will continue to bring about good results for some time to come. Later on selection will eliminate hens with undesirable genotypes. Only good hens will remain (over 180 eggs annually) and the effect of epistasis will vanish.

Table XV, analogous to Tables XII and XIII, shows 33 progenies on the farm of Chorzelów. Here the variance of variable t is normal for all fractions p (see the last row of Table XV) and for $p = 0,5$ the distribution of that variable is almost perfectly binomial (see Fig. 9). On this farm we encounter epistasis. Ordinary selection cannot give positive results here. It would only be possible to select two pure lines and produce a heterozygotic cross-breed, but even that method could not be expected to bring about a high laying capacity of the cross-breed (on the average 173,5 eggs a year).

In Figs. 7, 8 and 10 the distribution of the laying capacity of hens from the three cases considered is compared with the normal distribution (the histograms of the corresponding normal distributions are drawn with a dotted line). It will be seen that the distribution of the laying capacity of hens in Chorzelów comes nearest to the normal distribution and that of the hens which had been undergoing selection for 4 years in Czechnica shows the greatest divergence from the normal distribution.

The test based on progenies totalling 5 daughters each disregards a large amount of information contained in the empirical material (see for instance the laying capacities of daughters in Table X which are not underlined). The search for a test which would take all the information into consideration, i.e. which would handle progenies of various sizes, has failed to produce results. It has been found that the distributions of the corresponding variables t are extended too much and their variances are greater than normal in every case. *The test for comparing empirical and theoretical variances* has proved of too small power. Table XVII shows an analysis of various sizes of the progeny k . The smaller is the accepted size k , the worse will be the estimation of the variance of variable t . The greater the size k of progeny, the smaller will be the number n of progenies of that size. The size $k = 5$ seems to be optimal for the material under consideration.
