POSTĘPY NAUK ROLNICZYCH
nr 3 (99) 1966

BOLESŁAW ŻUK
Zakład Genetyki i Selekcji Zwierząt WSR we Wrocławiu
Kierownik: doc. dr B. Nowicki

METODA WSPÓŁCZYNNIKÓW ŚCIEŻEK WRIGHTA I JEJ NIEKTÓRE ZASTOSOWANIA

Wstęp

Od kilku lat w Polsce notuje się wyraźny wzrost zainteresowania metodami genetyki populacji. Coraz częściej stosuje się je w hodowli zwierząt. Wydaje się, że jeszcze szybszy wzrost zastosowań krępowany jest częściowo brakiem podręczników i opracowań wyjaśniających dlaczego w danym przypadku taką a nie inną metodę się stosuje.

Teoria współczynników ścieżek Wrighta (1921, 1934) leży u podstaw większości, jeśli nie wszystkich metod genetyki populacji stosowanych w hodowli zwierząt. Celem niniejszego artykułu jest omówienie tej teorii z podaniem podstaw matematycznych. W końcowej części podanych jest kilka przykładów rozwiązywania różnych problemów genetycznych.

Przy opracowaniu artykułu, oprócz prac cytowanych w tekście, wykorzystany był także podręcznik C. C. Li (1955).

Opis metody i zasady stosowania

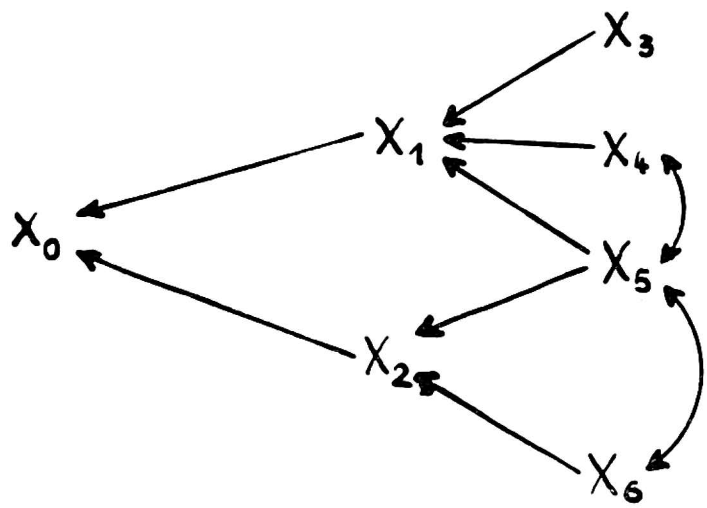
Często zakłada się z góry albo stwierdza doświadczalnie, że jakaś jedna zmienna jest skutkiem działania kilku czynników. Np. zakłada się, że wydajność produkcyjna zwierzęcia jest skutkiem działania czynnika genetycznego i czynnika środowiskowego. W wielu wypadkach istnieje także oczywisty matematyczny związek między zmiennymi; np. końcowy ciężar zwierzęcia, X, jest sumą (skutkiem) ciężaru przy urodzeniu, A i przyrostu, B:

$$X = A + B$$

Zmienne takie, jak w powyższych przykładach czynnik genetyczny, czynnik środowiskowy oraz ciężar początkowy i przyrost, nazywać będziemy p r z y c z y n a m i, a wynik ich działań (wydajność i ciężar końcowy) — ich s k u t k a m i.

Teoria współczynników ścieżek zajmuje się przedstawieniem związków i zależności między zespołem zmiennych przyczynowych i skutków. Używa się jej najczęściej do określania zależności (korelacji) między dwiema zmiennymi, z których jedna jest przyczyną drugiej, bądź też, gdy obie są skutkami działań tych samych przyczyn.

Metoda współczynnika ścieżki opiera się na konstrukcji jakościowego diagramu przedstawiającego zależności między pewną grupą zmiennych. Każda zmienna przedstawiona jest albo jako całkowicie określona przez pewne inne zmienne (które z kolei mogą być podobnie określone), albo jako czynnik końcowy [1] w tym diagramie. Zależności przyczynowe oznaczone są strzałkami. Czynniki końcowe powinny być połączone strzałkami z dwoma ostrzami. Strzałki te mają wskazywać możliwe korelacje przez jeszcze odleglejsze czynniki nie przedstawione w diagramie. Takie połączenia można opuścić, jeśli można bezpiecznie założyć, że czynniki końcowe nie są między sobą skorelowane. Diagram musi być kompletny, tzn. muszą w nim występować wszystkie przyczyny danego skutku. Rys. 1 przedstawia taki diagram. Na nim zmienna X_0 jest określona



Rys. 1

przez swoje przyczyny X_1 i X_2. Z kolei X_1 jest skutkiem dla przyczyn X_3, X_4, X_5, a X_2 jest skutkiem dla przyczyn X_5 i X_6. X_5 jest wspólną przyczyną dla X_1 i X_2 i powoduje część korelacji między X_1 i X_2. Zmienne X_3, X_4, X_5, X_6 są czynnikami końcowymi na tym diagramie. Między X_4 i X_5 oraz X_5 i X_6 istnieje korelacja, natomiast między pozostałymi korelacje takie nie występują.

Łańcuch strzałek łączący jakieś dwie zmienne nazywać będziemy *złożoną ścieżką* łączącą te zmienne, a pojedynczą strzałkę w takim łańcuchu — *ścieżką elementarną*. Każdej elementarnej ścieżce między przyczyną a skutkiem przyporządkowana jest pewna licz-

[1] Termin „czynnik końcowy" nie oznacza przyczyny pierwotnej, od której wywodzi się cały łańcuch zależności przyczynowych. Odnosi się on jedynie do konkretnego diagramu i oznacza, że przed czynnikiem końcowym diagram nie zawiera żadnych innych przyczyn.

ba zwana **w s p ó ł c z y n n i k i e m ś c i e ż k i**. Dokładna definicja współczynnika ścieżki zostanie podana później. Ścieżce złożonej, składającej się z łańcucha ścieżek elementarnych, odpowiada złożony współczynnik ścieżki. Współczynnik ścieżki od przyczyny X_j do skutku X_i oznaczać będziemy przez p_{ij}. Np. ścieżce od X_1 do X_0 ($X_0 \leftarrow X_1$) odpowiada współczynnik p_{01}. Złożony współczynnik ścieżki, odpowiadający np. ścieżce $X_0 \leftarrow X_1 \leftarrow X_4$, oznaczymy przez p_{014}. Ścieżce dwukierunkowej, np. $X_0 \leftarrow X_1 \leftarrow X_5 \rightarrow X_2$, odpowiadać będzie złożony współczynnik $p_{01(5)2}$ (nawias zaznacza najdalszy współczynnik w tym łańcuchu). Wskaźnik korelacji między końcowymi czynnikami, np. X_4 i X_5, oznaczać będziemy przez r_{45}.

Podane zostaną teraz zasady stosowane przy posługiwaniu się metodą współczynnika ścieżki.

1. Złożony współczynnik ścieżki jest równy iloczynowi współczynników elementarnych ścieżek wchodzących w skład całej ścieżki. (Np. rys. 1):

dla ścieżki $X_0 \leftarrow X_1 \leftarrow X_3 \quad p_{013} = p_{01}\, p_{13}$
dla ścieżki $X_0 \leftarrow X_1 \leftarrow X_5 \rightarrow X_2 \quad p_{01(5)2} = p_{01}\, p_{15}\, p_{25}$
dla ścieżki $X_0 \leftarrow X_2 \leftarrow X_6 \longleftrightarrow X_5 \rightarrow X_1 \; p_{02(65)1} = p_{02}\, p_{26}\, r_{56}\, p_{15}$

2. Wskaźnik koleracji między dwiema zmiennymi w diagramie jest równy sumie złożonych współczynników wszystkich ścieżek łączących te dwie zmienne. Np. zmienne X_1 i X_2 łączą następujące ścieżki:

$X_1 \leftarrow X_4 \longleftrightarrow X_5 \rightarrow X_2$ — współczynnik $p_{1(45)2}$
$X_1 \leftarrow X_5 \longleftrightarrow X_6 \rightarrow X_2$ — współczynnik $p_{1(56)2}$
$X_1 \leftarrow X_5 \rightarrow X_2$ — współczynnik $p_{1(5)2}$

więc

$$r_{12} = p_{1(45)2} + p_{1(56)2} + p_{1(5)2} = p_{14}\, r_{45} p_{25} + p_{15}\, r_{56}\, p_{26} + p_{15} p_{25}$$

3. Przy wyszukiwaniu właściwych ścieżek obowiązuje zasada: „nigdy w tył po pójściu naprzód". Zgodnie więc z tą zasadą w poprzednim przykładzie nie uwzględniliśmy ścieżki $X_1 \rightarrow X_0 \leftarrow X_2$, gdyż wtedy właśnie trzeba byłoby od X_1 iść naprzód (zgodnie z kierunkiem strzałki) do X_0 a następnie cofać się. Zasada „iść naprzód po pójściu w tył" jest prawidłowa.

Z zasady tej wynika zasada następna:

4. W skład ścieżki łączącej dwie zmienne może wchodzić co najwyżej jedna strzałka z dwoma ostrzami. A więc w powyższym przykładzie nie można uwzględniać ścieżki $X_1 \leftarrow X_4 \longleftrightarrow X_5 \longleftrightarrow X_6 \rightarrow X_2$ bo przechodząc od X_4 do X_5 idziemy częściowo naprzód, a następnie częściowo idziemy do tyłu.

5. Przy wyborze ścieżek nie wolno przechodzić dwa razy przez tę samą zmienną. Np. nie można uważać za ścieżkę łączącą ścieżki X_0 i X_1 ścieżki $X_0 \leftarrow X_1 \leftarrow X_4 \longleftrightarrow X_5 \rightarrow X_1$, bo przechodząc przez X_1 przechodzi się znowu do X_1.

Uzasadnienie metody

Na wstępie podamy trochę wiadomości ze statystyki, których znajomość jest niezbędna do zrozumienia całej teorii.

Dwie zmienne nazywamy niezależnymi, jeśli zmiany wartości jednej z nich nie zależą od zmiany wartości drugiej. Jeśli taka zależność istnieje, zmienne nazywamy zależnymi. Miarą takiej współzależności dwóch zmiennych jest wskaźnik korelacji. Wskaźnik korelacji, r_{AB} między dwiema zmiennymi A i B definiuje się następująco:

$$r_{AB} = \frac{\sigma_{AB}}{\sigma_A \, \sigma_B}$$

$\sigma_{AB} = \frac{1}{n} \Sigma(A-\bar{A})\,(B-\bar{B})$ nazywa się kowariancją zmiennych A i B

a $\sigma_A = \sqrt{\frac{1}{n} \Sigma(A-\bar{A})^2}$ i $\sigma_B = \sqrt{\frac{1}{n} \Sigma(B-\bar{B})^2}$ nazywają się standardowymi odchyleniami zmiennych A i B. Kwadrat standardowego odchylenia nazywa się wariancją. Symbole $\bar{A}$, $\bar{B}$ oznaczają wartości średnie A i B. Jeśli A i B są niezależne wówczas: $\sigma_{AB} = 0$ i $r_{AB} = 0$. Gdy $r_{AB} = 0$ zmienne takie nazywamy nieskorelowanymi.

Rozpatrzmy zmienną, która jest sumą dwóch zmiennych:

$$X = A + B$$

wariancja X, σ_x^2, jest równa

$$\sigma_x^2 = \frac{1}{n} \Sigma (X-\bar{X})^2 = \frac{1}{n} \Sigma[(A+B)-(\bar{A}+\bar{B})]^2 \quad (\text{ponieważ } \bar{X} = \bar{A}+\bar{B})$$

$$= \frac{1}{n} \Sigma[(A-\bar{A})+(B-\bar{B})]^2 = \frac{1}{n} \Sigma[(A-\bar{A})^2+2(A-\bar{A})(B-\bar{B})+(B-\bar{B})^2]$$

$$= \frac{1}{n} \Sigma(A-\bar{A})^2 + 2 \cdot \frac{1}{n} \Sigma(A-\bar{A})(B-\bar{B}) + \frac{1}{n} \Sigma(B-\bar{B})^2$$

$$= \sigma_A^2 + 2\sigma_{AB} + \sigma_B^2$$

Gdy A i B są niezależne, wówczas $\sigma_{AB} = 0$ i $\sigma_x^2 = \sigma_A + \sigma_B^2$ (1)

Obliczmy teraz korelację między X i A.

$$r_{XA} = \frac{\sigma_{XA}}{\sigma_X \, \sigma_A}$$

$$\sigma_{XA} = \frac{1}{n}\Sigma(X-\bar{X})\,(A-\bar{A}) = \frac{1}{n}\Sigma\,[(A-\bar{A})+(B-\bar{B})]\;(A-\bar{A}) =$$

$$= \frac{1}{n}\Sigma\,(A-\bar{A})^2 + \frac{1}{n}\,(A-\bar{A})\;(B-\bar{B}) = \sigma^2_A + \sigma_{AB}$$

Gdy A i B są niezależne, wówczas:

$$\sigma_{XA} = \sigma^2_A \tag{2}$$

i

$$r_{XA} = \frac{\sigma^2_A}{\sigma_X\;\sigma_A} = \frac{\sigma_A}{\sigma_X} \tag{3}$$

Gdybyśmy mieli dwie zmienne X i Y, przy czym $X = A + B$ i $Y = C + D$, wówczas:

$$\sigma_{XY} = \frac{1}{n}\,\Sigma(X-\bar{X})\,(Y-\bar{Y}) = \frac{1}{n}\,\Sigma[(A-\bar{A})+(B-\bar{B})]\;[(C-\bar{C}\quad I-\bar{I} =$$

$$= \frac{1}{n}\,\Sigma\,(A-\bar{A})\;(C-\bar{C}) + \frac{1}{n}\,\Sigma(A-\bar{A})\,(D-\bar{D}) + \frac{1}{n}\,\Sigma(B-\bar{B})\;(C-\bar{C})$$

$$+ \frac{1}{n}\,\Sigma(B-\bar{B})\;(D-\bar{D}) = \sigma_{AC} + \sigma_{AD} + \sigma_{BC} + \sigma_{BD}$$

Gdyby zmienne X i Y byłyby skutkami tych samych przyczyn, tzn. gdyby $A = C$ i $B = D$, wówczas:

$$\sigma_{AC} = \sigma^2_A,\; \sigma_{BD} = \sigma^2_B,\; \sigma_{AD} = \sigma_{BC} = \sigma_{AB} \text{ i otrzymamy}$$

$$\sigma_{XY} = \sigma^2_A + \sigma^2_B + 2\,\sigma_{AB} \tag{4}$$

Mamy pewną zmienną V_o. Jest ona skutkiem działania zmiennych przyczynowych $V_1, \ldots, V_n$.

Załóżmy dla uproszczenia, że średnie wszystkich zmiennych są równe 0. Zależność V_0 od $V_1, \ldots, V_n$ można przedstawić r ó w n a n i e m r e g r e s j i:

$$V_0 = b_{0,1}\;V_1 + b_{0,2}\;V_2 + \;.\;.\;.\; + b_{0,n}\;V_n \tag{5}$$

Współczynniki $b_{0,1}$, $b_{0,2}$, ,,, , $b_{0,n}$ nazywają się w s p ó ł c z y n n i k a m i r e g r e s j i c z ą s t k o w e j.

Współczynnik $b_{o,i}$ mierzy bezpośredni wkład zmiennej V_i w zmienną V_o. Bezpośrednią zależność między V_i i V_o mierzy w s k a ź n i k k o r e l a c j i c z ą s t k o w e j $r_{o,i}$. Gdy standardowe odchylenia wszystkich zmiennych V_0, V_1, ... V_n są sobie równe, wówczas $b_{o,i} = r_{o,i}$.

Równość przedstawiona przez (5) ma charakter s t a t y s t y c z n y, tzn. mierząc zmienne V_o, V_1, ..., V_n i podstawiając pomiary do równania (5) nie musi się otrzymać, i na ogół nie otrzymuje się, ścisłej równości lewej i prawej strony. Równanie regresji gwarantuje jedynie, że zgodność będzie możliwie najlepsza. Brak całkowitej zgodności obu stron równania (5) wynika stąd, że wśród zmiennych $V_1, ..., V_n$ nie występują wszystkie przyczyny skutku V_o. Czynniki te mogą być po prostu niemożliwe do zmierzenia. Oznaczmy wszystkie takie czynniki przez V_u. Uwzględniając jeszcze tę przyczynę, w miejscu równania (5) otrzyma się równanie:

$$V_o = C_{01} V_1 + C_{02} V_2 + ... + C_{0n} V_n + C_{ou} V_u \quad (6)$$

Tutaj otrzymujemy już równość matematyczą, tzn. gdyby nam się udało zmierzyć oprócz zmiennych $V_0, V_1, ..., V_n$ także zmienną V_u, wówczas wstawiając te pomiary do równania (6) otrzymalibyśmy dokładną równość lewej i prawej strony. Współczynniki C_{01}, itd. są więc typu współczynników regresji cząstkowej, ale nie są z nimi równoznaczne. Nazywają się one współczynnikami regresji ścieżkowej (path regression coefficient, Wright, 1921). „Współczynnik regresji ścieżkowej C_{01}, mierzy konkretny wkład V_1 w V_o z punktu widzenia przedstawionego w diagramie. Jeśli diagram przedstawia prawidłowo związki przyczynowe, regresja ścieżkowa mierzy ten wkład w sensie absolutnym i jej wartość może być użyta w analizie innych populacji" (Wright, 1960).

Gdy σ_0, σ_1, itd. oznaczają standardowe odchylenia zmiennych V_0, V_1, itd., zmienne $X_o = \frac{V_0}{\sigma_o}, X_1 \frac{V_1}{\sigma_1}$ itd. mają standardowe odchylenia równe 1. Zmienne X_0, X_1 itd. nazywają się zmiennymi standaryzowanymi. Podstawiając $V_o = X_o\sigma_o$, $V_1 = X_1\sigma_1$ itd. do równania (6) i dzieląc obie strony przez σ_o otrzymujemy:

$$X_0 = C_{01} \frac{\sigma_1}{\sigma_0} X_1 + C_{02} \frac{\sigma_2}{\sigma_0} X_2 + \ldots + C_{0n} \frac{\sigma_n}{\sigma_0} X_n + C_{0u} \frac{\sigma_u}{\sigma_0} X_u \quad (7)$$

Wielkość

$$p_{0i} = C_{oi} \frac{\sigma_i}{\sigma_o} \quad (8)$$

nazywa się współczynnikiem ścieżki (path coefficient) od przyczyny V_i do skutku V_o i mierzy bezpośrednią zależność skutku V_o od przyczyny V_i. Współczynniki ścieżki w równaniu (7) są współczynnikami regresji ścieżkowej, można zatem powiedzieć, że współczynnik ścieżki jest „standaryzowanym" współczynnikiem regresji ścieżkowej.

Współczynnik ścieżki jest analogiczny do wskaźnika korelacji cząstkowej i różni się od niego tym, że dotyczy systemu zmiennych, z których jedna jest skutkiem, a pozostałe stanowią pełny zespół przyczyn. Poza tym współczynnik ścieżki ma kierunek, tzn. odpowiada on zmiennym, z których jedna jest przyczyną, a druga skutkiem. Gdyby na innym diagramie zmienne te zamieniły się rolami: tzn. „przyczyna" stałaby się „skutkiem", a „skutek" — „przyczyną", wówczas dla nowego kierunku ścieżki otrzymamy na ogół inny współczynnik ścieżki.

Współczynnik regresji cząstkowej C_{0i} mierzy bezpośredni układ zmiennej V_i w zmienność V_0, zatem iloczyn $C_{0i} \cdot \sigma_1$ oznacza tę część standardowego odchylenia skutku V_0, którą wywołuje V_i. Oznaczmy tę część przez $\sigma_{0:1}$, wtedy (8) można zapisać tak:

$$p_{0i} = \frac{\sigma_{0:1}}{\sigma_0} \tag{9}$$

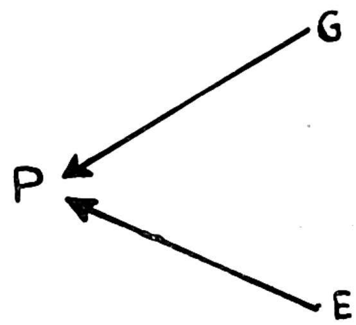
Otrzymujemy jeszcze jedną, równoważną definicję współczynnika ścieżki; współczynnik ścieżki jest to liczba mówiąca, jaka część ogólnego standardowego odchylenia skutku V_0 stanowi standardowe odchylenie wywołane wpływem przyczyny V_i.

Przed przystąpieniem do omówienia własności współczynników ścieżek zdefiniujemy jeszcze pewną wielkość ściśle się z nim wiążącą. System zmiennych $V_1, ..., V_u$ całkowicie determinuje zmienną V_0. W tej ogólnej determinacji różne zmienne mają różny wkład. W s k a ź n i k i e m d e t e r m i n a c j i skutku V_0 przez przyczynę V_i, d_{0i} nazywa się liczba mówiąca jaką część ogólnej zmienności zmiennej V_0 stanowi zmienność wywołana przez V_i.
Oczywisty jest związek:

$$d_{01} + d_{02} + ... + d_{on} + d_{ou} = 1 \tag{10}$$

Rozpatrzmy teraz kilka elementarnych schematów przyczynowych. Na takie elementarne schematy można w zasadzie rozłożyć każdy złożony schemat, z jakim możemy się spotkać w genetyce.

P r z y c z y n y n i e z a l e ż n e. Niech E i G będą nieskorelowanymi zmiennymi całkowicie determinującymi zmienną P. Przedstawia to rys. 2.



Rys. 2

Gdy G i E są niezależne ($r_{GE} = 0$) $\sigma_p^2 = \sigma_G^2 + \sigma_E^2$ [patrz wzór (1)], a więc σ_G^2 mierzy tę część ogolnej zmienności P, która jest bezpośrednio determinowana przez G. Zatem wskaźnik determinacji P przez G jest równy $d_{PG} = \frac{\sigma_G^2}{\sigma_P^2}$. Podobnie $d_{PE} = \frac{\sigma_E^2}{\sigma_P^2}$ Równość (9) jest spełniona bo:

$$d_{PG} + d_{PE} = \frac{\sigma_G^2}{\sigma_P^2} + \frac{\sigma_E^2}{\sigma_P^2} = \frac{\sigma_G^2 + \sigma_E^2}{\sigma_P^2} = \frac{\sigma_P^2}{\sigma_P^2} = 1$$

Standardowym odchyleniem P jest σ_P, a jego częścią, spowodowaną wpływem zmiennej G jest σ_G. Zatem, wobec (9):

$$p_{PG} = \frac{\sigma_G}{\sigma_P}$$

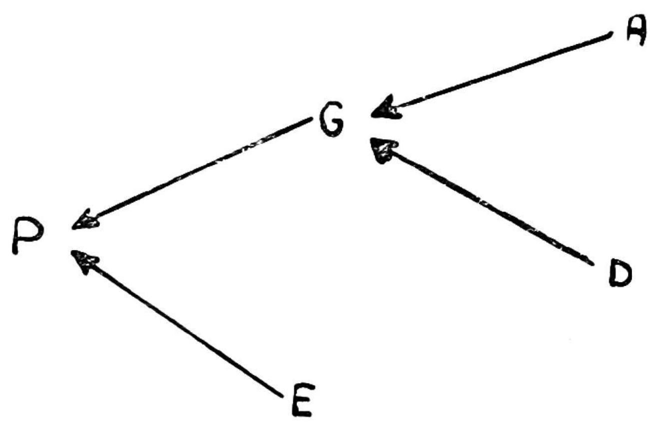
Podobnie: $$p_{PE} = \sigma_E/\sigma_P$$

Porównując uzyskane rezultaty z (3) (w którym to wyrażeniu A trzeba zamienić na G lub E, a X na P) stwierdzamy dla przyczyn niezależnych:

$$p_{PG} = \sqrt{d_{PG}} = r_{PG} = \frac{\sigma_G}{\sigma_P} \tag{11}$$

Podobne związki w podobny sposób można otrzymać, gdy występuje więcej niż dwie niezależne zmienne przyczynowe.

Łańcuch zmiennych niezależnych. W jakimś systemie przyczynowym któraś z przyczyn może być skutkiem działania jakichś innych przyczyn. Rys. 3 przedstawia sytuację, gdzie przyczyna (dla P)



Rys. 3

G jest już skutkiem działania wcześniejszych przyczyn A i D („wcześniejszy" niekoniecznie musi oznaczać wcześniejszy w czasie). Zakładamy, że A i D, podobnie jak G i E, są niezależne.

Mamy tutaj $\sigma_P^2 = \sigma_G^2 + \sigma_E^2$, ale ponieważ $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$ a więc $\sigma_P^2 = \sigma_A^2 + \sigma_D^2 + \sigma_E^2$. Rozumując jak poprzednio otrzymujemy: $p_{PG} = \frac{\sigma_G}{\sigma_P}$, $p_{GA} = \frac{\sigma_A}{\sigma_G}$

Mnożąc p_{PG} przez p_{GA} otrzymujemy:

$$p_{PG} \cdot p_{GA} = \frac{\sigma_G}{\sigma_P} \cdot \frac{\sigma_A}{\sigma_G} = \frac{\sigma_A}{\sigma_P} = p_{PGA}$$

(p_{PGA} — złożony współczynnik ścieżki $P \leftarrow G \leftarrow A$). Podobnie można wyprowadzić następne związki:

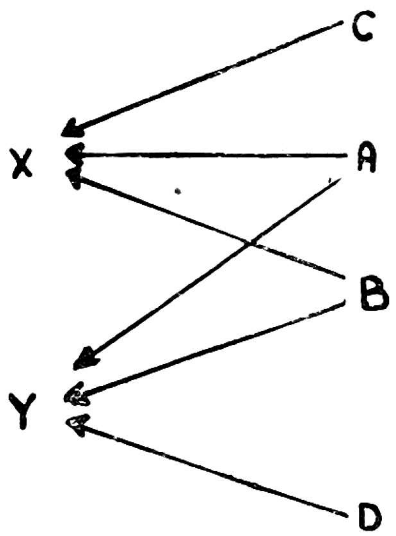
$$d_{PA} = d_{PG} \cdot d_{GA}$$
$$r_{PA} = r_{PG} \cdot r_{GA}$$

Gdyby któraś z przyczyn, np. A, miała jeszcze jedną przyczynę, np. B, wówczas złożony współczynnik ścieżki p_{PGAB} byłby równy

$$p_{PGAB} = p_{PG} \cdot p_{GA} \cdot p_{AB} \tag{12}$$

Podobnie będzie z wskaźnikami determinacji i korelacji. Uogólnienie na większą ilość zmiennych w łańcuchu jest oczywiste. Ze związku (12) wynika zasada 1 w rozdziale 2.

W s p ó l n e   p r z y c z y n y   i   k o r e l a c j a. Dwa różne skutki mogą mieć te same przyczyny. Taką sytuację mamy na rys. 4, gdzie A i B



Rys. 4

są wspólnymi przyczynami zmiennych X i Y. Poza nimi przyczyną X jest jeszcze C, a D — przyczyną Y. Wszystkie przyczyny są między sobą niezależne. Wspólne przyczyny A i B powodują, że zmienne X i Y są skorelowane między sobą. Jaki jest wskaźnik korelacji między nimi? Z definicji:

$$r_{XY} = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y}$$

Wobec niezależności przyczyn $\sigma_{XY} = \sigma_A^2 + \sigma_B^2$, zatem

$$r_{XY} = \frac{\sigma_A^2 + \sigma_B^2}{\sigma_X \, \sigma_Y} = \frac{\sigma_A^2}{\sigma_X \, \sigma_Y} + \frac{\sigma_B^2}{\sigma_X \, \sigma_Y} = \frac{\sigma_A}{\sigma_X} \cdot \frac{\sigma_A}{\sigma_Y} + \frac{\sigma_B}{\sigma_X} \cdot \frac{\sigma_B}{\sigma_Y} =$$

$= p_{XA} \cdot p_{YA} + p_{XB} \cdot p_{YB}$ i, wobec zasady 1 rozdziału 2

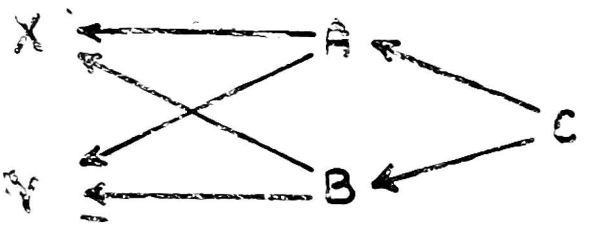
$$r_{XY} = p_{X(A)Y} + p_{X(B)Y}$$

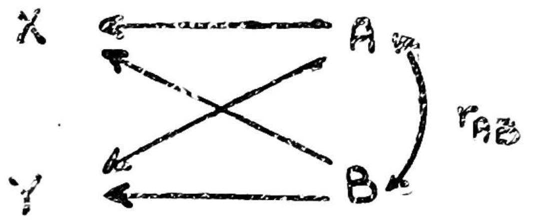
($p_{X(A)Y}$ — złożony współczynnik ścieżki odpowiadający ścieżce $X \leftarrow A \rightarrow Y$ i analogicznie $p_{X(B)Y}$).

S k o r e l o w a n e p r z y c z y n y. Na rys. 5 przedstawiona jest sytuacja, gdy bezpośrednimi przyczynami skutków X i Y są A i B. Z kolei C jest przyczyną dla A i B, zatem przyczyny A i B są skorelowane. Wobec (4):

$$r_{XY} = \frac{\sigma_A^2 + \sigma_B^2 + 2\,\sigma_{AB}}{\sigma_X\,\sigma_Y} = \frac{\sigma_A}{\sigma_X} \cdot \frac{\sigma_A}{\sigma_Y} + \frac{\sigma_B}{\sigma_X} \cdot \frac{\sigma_B}{\sigma_Y} + 2\,\frac{\sigma_{AB}}{\sigma_X\,\sigma_Y};$$

$$2\,\frac{\sigma_{AB}}{\sigma_X\,\sigma_Y} = 2\,\frac{\sigma_{AB}}{\sigma_X\,\sigma_Y} \cdot \frac{\sigma_A}{\sigma_A} \cdot \frac{\sigma_B}{\sigma_B} = 2\,\frac{\sigma_{AB}}{\sigma_A\,\sigma_B} \cdot \frac{\sigma_A}{\sigma_X} \cdot \frac{\sigma_B}{\sigma_Y} = 2 r_{AB} \cdot \frac{\sigma_A}{\sigma_X}\,\frac{\sigma_B}{\sigma_Y}$$





Rys. 5

Ponieważ $\frac{\sigma_A}{\sigma_X} = p_{AX}$ itd., otrzymujemy:

$$r_{XY} = p_{XA}\,p_{YA} + p_{XB}\,p_{YB} + 2\,r_{AB}\,p_{XA}\,p_{XB} \tag{12}$$

Skorelowanie przyczyn A i B przez wspólną ich przyczynę C wprowadza do wzoru na r_{XY} korelację między A i B. Schemat z rys. 5 można uprościć oznaczając skorelowanie strzałką z dwoma ostrzami (rys. 5a).

Teraz (12) można przedstawić przy pomocy złożonych współczynników ścieżki: $p_{XA}; p_{YA}$ jest złożonym współczynnikiem dla ścieżki $X \leftarrow A \rightarrow Y$, $p_{XB}\,p_{YB}$ dla ścieżki $X \leftarrow B \rightarrow Y$, $r_{AB}\,p_{XA}\,p_{YB}$ dla ścieżki $X \leftarrow A \longleftrightarrow B \rightarrow Y$ i $r_{AB}\,p_{XA}\,p_{YB}$ dla ścieżki $X \leftarrow B \longleftrightarrow A \rightarrow Y$. Zatem

$$r_{XY} = p_{X(A)Y} + p_{X(B)Y} + {}_{X(BA)Y}$$

Stąd bierze się zasada 2 na str. 87.

Wróćmy jeszcze do rys. 2. Na nim zmienne G i E są niezależne, czyli $r_{GF} = 0$. Skutek P jest połączony pojedynczą ścieżką z przyczyną G, a także jest połączony pojedynczą ścieżką z przyczyną E. Ale nie można uważać, że G i E są połączone ścieżką przez P, gdyż wówczas korelacja między nimi byłaby równa iloczynowi odpowiednich ścieżek, czyli byłaby różna od 0, a z założenia mamy $r_{GE} = 0$. Stąd bierze się zasada 3 (str. 87).

Zasada 5 wynika z determinacji skutku przez przyczyny. Na rys. 1

zmienna X_0 jest całkowicie zdeterminowana przez przyczyny X_1, X_2. Jest ona determinowana także przez dalsze przyczyny, np. przez X_3, X_4, X_5, ale determinacja przez te przyczyny tkwi już w determinacji przez X_1.

Przykłady

W wielu problemach genetycznych panujące zależności można przedstawić przy pomocy schematu przyczynowego. Przy rozwiązaniu takich problemów pomocna się stała teoria współczynników ścieżek. I tak umożliwiła ona ocenę spokrewnienia między zwierzętami. Umożliwiła ona zmierzenie stopnia nasilenia chowu w pokrewieństwie. Najważniejsza praca w hodowli zwierząt — selekcja — ma sens tylko wtedy, gdy na daną cechę ma wpływ czynnik genetyczny. Współczynnik ścieżki znowu pozwala na zmierzenie tego wpływu. Selekcji dokonuje się na podstawie wartości hodowlanych zwierząt, które szacuje się na podstawie produkcji własnej i krewnych. Potrzebna więc jest znajomość zależności między genotypami i fenotypami krewnych. Pomiar tych zależności także stał się możliwy dzięki teorii współczynników ścieżek. Przy selekcji na kilka cech dochodzą jeszcze zależności między cechami. Tu także wykorzystuje się współczynniki ścieżek.

Już tych kilka przykładów wskazuje na znakomitą rolę teorii w hodowli zwierząt. Poniżej pokazane zostaną przykłady zastosowania metody współczynników ścieżek do rozwiązania kilku problemów genetycznych.

Przykład 1.

Rys. 6 przedstawia diagram zależności między genotypami rodziców i ich potomstwa. Korelacje między rodzicami pokazują strzałki o podwójnych ostrzach ze wskaźnikami korelacji m_1, m_2 itd. Współczynnik ścieżki między rodzicem a potomkiem oznaczony jest przez a. Jest on taki sam dla każdej ścieżki rodzic-potomek, ponieważ każdy rodzic genotyp swego potomka determinuje w tym samym stopniu.

Współczynnik ścieżki od rodzica do potomka jest równy $^1/_2$. Sposób jego obliczenia można znaleźć w podręczniku C.C. Li (1955), str. 172—175.

K o r e l a c j a m i ę d z y p e ł n y m r o d z e ń s t w e m. Na rys. 6 pełnym rodzeństwem są P_1 i P_2. Łączą je ścieżki:

a) $P_1 \leftarrow O_1 \rightarrow P_2$ z współczynnikiem $a \cdot a = a^2 = \frac{1}{4}$

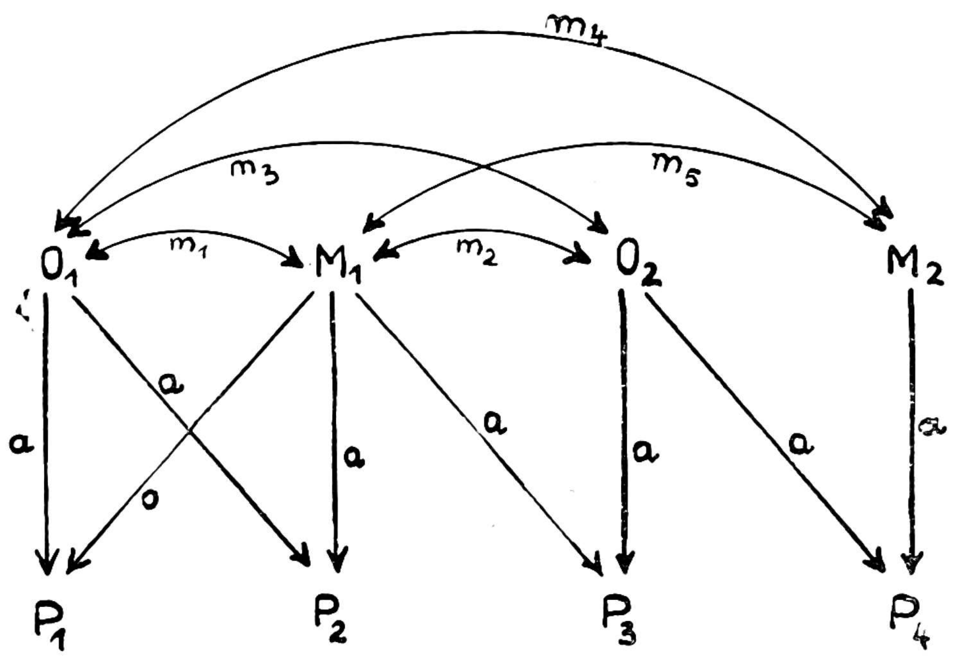
b) $P_1 \leftarrow M_1 \rightarrow P_2$ z współczynnikiem $a \cdot a = a^2 = \frac{1}{4}$

c) $P_1 \leftarrow O_1 \longleftrightarrow M_1 \rightarrow P_2$ współczynnikiem $a \cdot m_1 a = a^2 m_1 = \frac{1}{4} m_1$

d) $P_1 \leftarrow M_1 \longleftrightarrow O_1 \rightarrow P_2$ współczynnikiem $a \cdot m_1 a = a^2 m_1 = \frac{1}{4} m_1$

Zatem

$$r_{P_1P_2} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} m_1 + \frac{1}{4} m_1 = \frac{1 + m_1}{2} \tag{13}$$



Rys. 6

K o r e l a c j a m i ę d z y p ó ł r o d z e ń s t w e m. Półrodzeństwem są osobniki P_1 i P_3 (wspólny przodek M_1). Wyszukując ścieżki podobnie jak porzednio otrzymujemy:

$$r_{P_1P_3} = a^2 + a^2 m_1 + a^2 m_2 + a^2 m_3 = a^2 (1 + m_1 + m_2 + m_3) =$$
$$= \frac{1}{4} (1 + m_1 + m_2 + m_3) \tag{14}$$

K o r e l a c j a m i ę d z y n i e r o d z e ń s t w e m. Osobniki P_1 i P_4 są nierodzeństwem, ale korelację między nimi powoduje spokrewnienie rodziców. Mamy:

$$r_{P_1P_4} = a^2 m_2 + a^2 m_4 + a^2 m_3 + a^2 m_5 = \frac{1}{4} (m_2 + m_3 + m_4 + m_5)$$

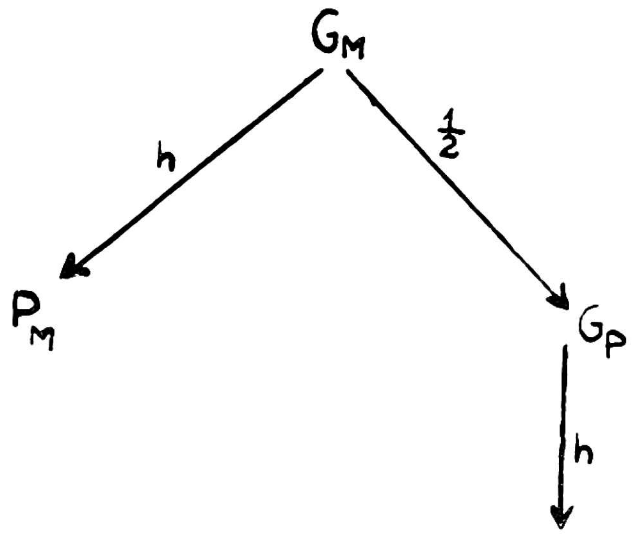
Wartości m_1, m_2, m_3, m_4, m_5 są równe współczynnikom pokrewieństwa między odpowiednimi osobnikami i można je obliczyć przy pomocy znanego wzoru.

Przykład 2.

Zakłada się, że fenotyp jakiegoś osobnika jest skutkiem działania czynnika genetycznego G i czynnika środowiskowego E. Przyjmuje się, że czynniki te nie są skorelowane. Sytuację taką przedstawia rys. 2. Wskaźnik determinacji P przez G w genetyce nazywa się odziedziczalnością (h^2) i mierzy on frakcję ogólnej zmienności spowodowanej wpływem czynnika genetycznego. Współczynnik ścieżki i wskaźnik korelacji G i P, wobec (11) jest równy:

$$p_{PG} = r_{PG} = \sqrt{h^2} = h \qquad (15)$$

Zależność między fenotypami rodzica i potomka. Rys. 7 przedstawia diagram pomocniczy.



Rys. 7

Współczynniki ścieżek $P_M \leftarrow G_M$ oraz $P_P \leftarrow G_P$ są równe h (wzór 15), a współczynnik ścieżki $G_P \leftarrow G_M$ — $\frac{1}{2}$ (przykład 1).

Zatem:

$$r_{P_M P_P} = h \cdot \frac{1}{2} \cdot h = \frac{1}{2} h^2$$

z równania tego można wyliczyć h^2:

$$h^2 = 2 r_{P_M P_P}$$

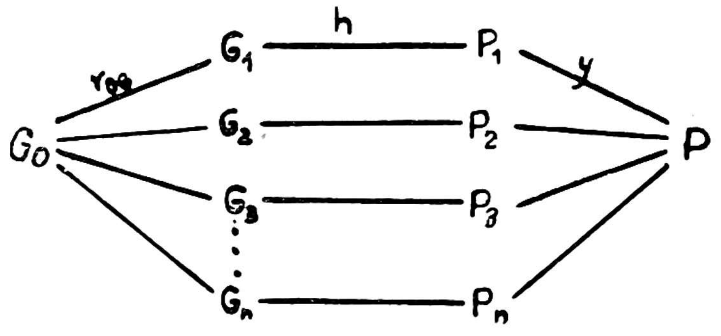
Na wzorze tym opiera się jedna z metod szacowania odziedziczalności, przy której odziedziczalność jest równa podwojonej korelacji między wydajnościami matki i córki.

Przykład 3.

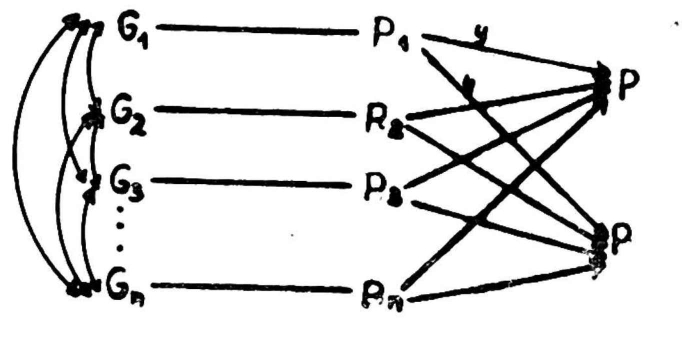
Zależność między genotypem rodzica i przeciętną fenotypów potomstwa. Rys. 8 przedstawia zależności przyczynowe między genotypem ojca G_o, a przeciętną fenotypów jego u potomków P. G_1, G_2, ..., G_n przedstawiają genotypy potomków, a P_1, P_2, ..., P_n — ich fenotypy. Współczynnik ścieżki od genotypu rodzica do genotypu potomka jest równy $\frac{1}{2}$, od genotypu potomka do fenotypu potomka — h [patrz (15)]. Współczynnik ścieżki od fenotypu potomka do przeciętnej P oznaczony został przez y.

Korelacja między G_o i P równa jest sumie złożonych współczynników ścieżek łączących G_o i P. Ścieżek takich jest n; dla każdej współczynnik jest równy $\frac{1}{2}$ hy. Zatem

$$r_{G_oP} = \frac{1}{2}hyn \tag{16}$$



Rys. 8



Rys. 8a

Trzeba obliczyć y. Do tego celu posłuży nam rys. 8a. Na rysunku tym obie wielkości P są tymi samymi wielkościami. Korelacja między nimi jest oczywiście równa 1. Z drugiej strony jest ona równa sumie współczynników wszystkich ścieżek łączących te dwie zmienne. Zakładamy, że korelacje między genotypami potomków są sobie równe i wynoszą m. Gdyby matki tych potomków nie były spokrewnione z ojcem, wówczas $m = \frac{1}{4}$ w przeciwnym wypadku m można obliczyć według wzoru (14) bądź według wzoru (13), jeśli wszyscy potomkowie mają tę samą matkę.

Przez P_1 przechodzą następujące ścieżki:

$P \leftarrow P_1 \rightarrow P$	współczynnik $y \cdot y = y^2$
$P \leftarrow P_1 \leftarrow G_1 \longleftrightarrow G_2 \rightarrow P_2 \rightarrow P$	współczynnik y^2h^2m
$P \leftarrow P_1 \leftarrow G_1 \longleftrightarrow G_n \rightarrow P_n \rightarrow P$	współczynnik y^2h^2m

W grupie tej jedna ścieżka ma współczynnik y^2, a n—1 ścieżek współczynnik y^2h^2m. Suma tych współczynników wynosi $y^2 + (n - 1)\, y^2h^2m = y^2 [1 + (n - 1)\, h^2m]$. Dla każdej z pozostałych grup ścieżek (jest ich wszystkich, łącznie z poprzednio wypisaną, n) sytuacja będzie identyczna, tzn. suma współczynników będzie równa: $y^2 [1 + (n - 1) h^2 m]$. Korelacja między P i P (równa 1) jest równa sumie współczynników ścieżek z wszystkich grup, tzn.:

$$r_{PP} = 1 = ny^2 [1 + (n - 1) h^2 m] = 1$$

Stąd:

$$y = \sqrt{\frac{1}{n[1 + (n - 1) h^2 m]}}$$

Podstawiając do wzoru (16) mamy:

$$r_{G_O P} = \frac{1}{2} hn \sqrt{\frac{1}{n[1 + (n - 1) h^2 m]}} = \frac{1}{2} h \sqrt{\frac{n}{1 + (n - 1) h^2 m}}$$

LITERATURA

1. Li C. C. 1955: Population genetics. Uniw. Chicago Press. Chicago.
2. Wright S. 1921: J. Agric. Res., 20, 557—585.
3. Wright S. 1934: Annals of Math. Stat., 5, 161—215.
4. Wright S. 1960.: Biometrics, 16, 189—202.