

WYKORZYSTANIE RÓŻNYCH ŹRÓDEŁ INFORMACJI W PREDYKCJI WARTOŚCI HODOWLANEJ

Dawid Słomian

- Model bazowy
- Predykcja wartości hodowlanej na podstawie wydajności własnej
- Predykcja wartości hodowlanej na podstawie wydajności potomstwa
- Predykcja wartości hodowlanej na podstawie wydajności krewnych
- Predykcja wartości hodowlanej cechy A na podstawie wydajności cechy B
- Indeks selekcyjny

Predykcja wartości hodowlanej

- Predykcja vs estymacja
- Estymator vs wartość prawdziwa
- Dokładność predykcji wartości hodowlanej zależy od dostępności źródeł informacji = rekordów danych
- Dostępność rekordów danych w czasie życia osobnika:
 - *Rekordy przodków*
 - *Własna wydajność*
 - *Wydajności potomstwa*

Model genetyczny

- fenotyp = środowisko + genotyp + błąd

$$y_{ij} = \beta_i + g_i + e_{ij}$$

- y_{ij} wydajność j osobnika i
- β_i efekt środowiskowy osobnika i np. płeć, rok urodzenia, stado
- g_i efekt genetyczny osobnika i = suma efektów addytywnych, dominacyjnych i epistatycznych wszystkich genów
- e_{ij} sumaryczny efekt innych czynników niegenetycznych warunkujących wydajność j osobnika i

Model hodowlany

- fenotyp = środowisko + efekt addytywnie genetyczny + błąd

$$y_{ij} = \beta_i + a_i + e_{ij}$$

- y_{ij} wydajność j osobnika i
- β_i efekt środowiskowy osobnika i np. płeć, rok urodzenia, stado
- a_i wartość hodowlana osobnika i = suma efektów addytywnych wszystkich genów
- e_{ij} sumaryczny efekt innych czynników środowiskowych oraz czynników genetycznych nieaddytywnych warunkujących wydajność j osobnika i

Model hodowlany → założenia

$$y_{ij} = \beta_i + a_i + e_{ij}$$

- $y \sim N(0, \sigma_y^2)$ wydajności są determinowane przez nieskończenie wiele niezależnych od siebie genów o efektach addytywnych → infinitesimal model
- Fisher (1918)
- $\sigma_y^2, \sigma_a^2, \sigma_e^2$, znane
- $cov(a, e) = 0$ $cov(e_{ij}, e_{ik}) = 0$

Model hodowlany → założenia

- $$a_i = \frac{1}{2} a_{si} + \frac{1}{2} a_{di} + m_i$$
- a_i wartość hodowlana osobnika i
- a_{si} wartość hodowlana ojca (sire) osobnika i
- a_{di} wartość hodowlana matki (dam) osobnika i
- m_i Mendelian sampling = odchylenie wartości hodowlanej osobnika i od średniej rodziców

Predykcja wartości hodowlanej na podstawie wydajności własnej

- Źródło informacji – pojedyncza wydajność osobnika i

$$\hat{a}_i = b(y_i - \beta_i)$$

- β_i średnia wartość fenotypowa osobników reprezentujących tą samą grupę czynników środowiskowych co osobnik i np. wspólne stado, płeć

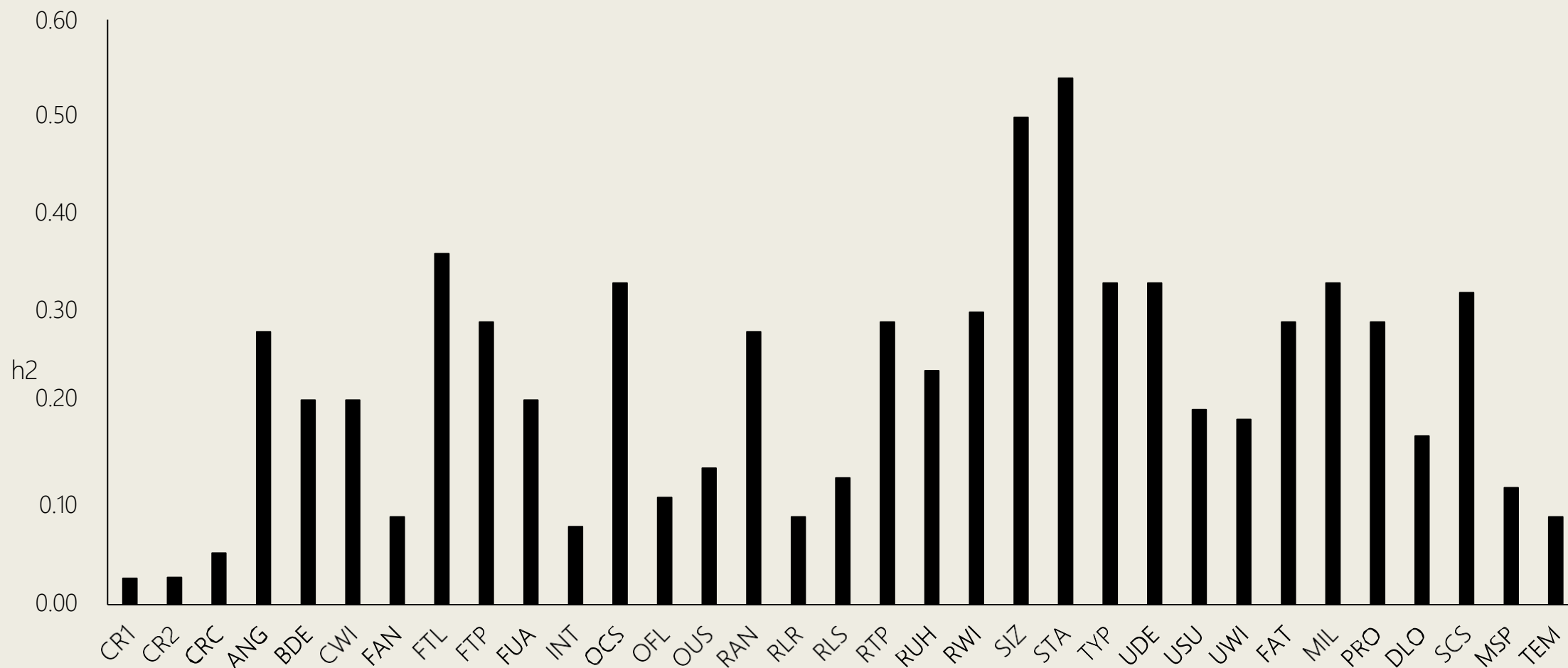
- b współczynnik regresji prawdziwej wartości hodowlanej (a) na wartość fenotypową (y)

$$b = \frac{\text{cov}(a, y)}{\text{var}(y)} = \frac{\text{cov}(a, a + e)}{\text{var}(y)} = \frac{\text{cov}(a, a) + \text{cov}(a, e)}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_y^2} = h^2$$

- h^2 odziedziczalność (heritability)

Predykcja wartości hodowlanej na podstawie wydajności własnej

■ Odziedziczalność wybranych cech rasy Polskiej Holsztyńsko-fryzyjskiej



Predykcja wartości hodowlanej na podstawie wydajności własnej

- Dokładność (accuracy) predykcji = korelacja pomiędzy kryterium selekcyjnym (y), a prawdziwą wartością hodowlaną (a)

$$r = \frac{\text{cov}(a, y)}{\sigma_a \sigma_y} = \frac{\sigma_a^2}{\sigma_a \sigma_y} = \frac{\sigma_a}{\sigma_y} = h$$

- Powtarzalność (repeatability) $r^2 = h^2$
- Oczekiwany postęp selekcyjny (expected response to selection)

$$R = ih^2\sigma_y$$

i intensywność selekcji, różnica pomiędzy osobnikami wybranymi do hodowli, a średnią całej populacji

Predykcja wartości hodowlanej na podstawie wydajności potomstwa

- Selekcja buhajów na podstawie wydajności córek
- Intensywność selekcji buhajów >> intensywność selekcji krów

Predykcja wartości hodowlanej na podstawie wydajności potomstwa

- Źródło informacji – pojedyncza wydajność osobnika i

$$\hat{a}_i = b(y_i - \beta_i)$$

- Źródło informacji – wydajności córek osobnika i

$$\hat{a}_i = b(\bar{y}_i - \beta_i)$$

- $\bar{y}_i$ średnia wydajność córek buhaja i

- β_i średnia wartość fenotypowa osobników reprezentujących tą samą grupę czynników środowiskowych co osobnik i np. wspólne stado, płeć

- b współczynnik regresji prawdziwej wartości hodowlanej (a) na wartość fenotypową (y)

Predykcja wartości hodowlanej na podstawie wydajności potomstwa

- b współczynnik regresji prawdziwej wartości hodowlanej (a) na wartość fenotypową (y)

- $$b = \frac{\text{cov}(a,y)}{\text{var}(y)} = \frac{\text{cov}(a,a+e)}{\text{var}(y)} = \frac{\sigma_a^2}{\sigma_y^2} = h^2$$

- $$b = \frac{\text{cov}(a,\bar{y})}{\text{var}(\bar{y})}$$

- $$\text{cov}(a,\bar{y}) = \text{cov}\left(a, \frac{1}{2}a_s + \frac{1}{2}a_d + \frac{\sum e_i}{n}\right) = \frac{1}{2}\text{cov}(a, a_s) = \frac{1}{2}\sigma_a^2 \quad \text{var}(\bar{y}) = \left(t + \frac{1-t}{n}\right)\sigma_y^2$$

- a_s wartość hodowlana buhaja i (ojciec potomstwa)
- a_d wartość hodowlana krów (matki potomstwa)
- t korelacja pomiędzy córkami buhaja i

Predykcja wartości hodowlanej na podstawie wydajności potomstwa

- b współczynnik regresji prawdziwej wartości hodowlanej (a) na wartość fenotypową (y)

$$b = \frac{cov(a, \bar{y})}{var(\bar{y})} = \frac{\frac{1}{2} \sigma_a^2}{\left(t + \frac{1-t}{n}\right) \sigma_y^2}$$

- Np. półsiostry, córki buhaja i : $t = \frac{1}{4} \frac{\sigma_a^2}{\sigma_y^2} = \frac{1}{4} h^2$

$$b = \frac{2n}{n+k}$$

$$k = \frac{4-h^2}{h^2}$$

zależny od:
odziedziczalności cechy
liczby córek
 $n \rightarrow \infty \quad b \rightarrow 2$

Predykcja wartości hodowlanej na podstawie wydajności potomstwa

- Dokładność (accuracy) predykcji = korelacja pomiędzy kryterium selekcyjnym (y), a prawdziwą wartością hodowlaną (a)

$$r = \frac{\text{cov}(a, y)}{\sigma_a \sigma_y} = \frac{\sigma_a^2}{\sigma_a \sigma_y} = \frac{\sigma_a}{\sigma_y} = h$$

$$r = \frac{\text{cov}(a, \bar{y})}{\sigma_a \sigma_{\bar{y}}} = \frac{\frac{1}{2} \sigma_a^2}{\sqrt{\sigma_a^2 \left(t + \frac{1-t}{n} \right) \sigma_y^2}}$$

Predykcja wartości hodowlanej na podstawie wydajności potomstwa

- Np. półsiostry, córki buhaja i : $t = \frac{1}{4} \frac{\sigma_a^2}{\sigma_y^2} = \frac{1}{4} h^2$

$$r = \frac{\text{cov}(\mathbf{a}, \bar{\mathbf{y}})}{\sigma_a \sigma_{\bar{y}}} = \frac{\frac{1}{2} \sigma_a^2}{\sqrt{\sigma_a^2 \left(t + \frac{1-t}{n} \right) \sigma_y^2}}$$

$$= \frac{\frac{1}{2} h^2 \sigma_y^2}{\sqrt{h^2 \sigma_y^2 \left(\frac{1}{4} h^2 + \frac{1 - \frac{1}{4} h^2}{n} \right)}} = \sqrt{\frac{n}{n + k}}$$

zależna od:
odziedziczalności cechy
liczby córek
 $n \rightarrow \infty \quad r \rightarrow 1$

Predykcja wartości hodowlanej na podstawie wydajności potomstwa

- Powtarzalność (repeatability)

$$r^2 = \frac{n}{n + k}$$

Predykcja wartości hodowlanych na podstawie wydajności krewnych

- Źródło informacji – pojedyncza wydajność osobnika i

$$\hat{a}_i = b(y_i - \beta_i)$$

- Źródło informacji – wydajności córek osobnika i

$$\hat{a}_i = b(\bar{y}_i - \beta_i)$$

- Źródło informacji – wartości hodowlane rodziców osobnika i

$$\hat{a}_i = \frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d$$

Predykcja wartości hodowlanych na podstawie wydajności krewnych

- Dokładność (accuracy) predykcji = korelacja pomiędzy kryterium selekcyjnym (y), a prawdziwą wartością hodowlaną (a)

$$r = \frac{\text{cov}(a, y)}{\sigma_a \sigma_y} = \frac{\sigma_a^2}{\sigma_a \sigma_y} = \frac{\sigma_a}{\sigma_y} = h$$

$$r = \frac{\text{cov}(a, \bar{y})}{\sigma_a \sigma_{\bar{y}}} = \sqrt{\frac{n}{n+k}}$$

$$r = \frac{\text{cov}(a, f)}{\sigma_a \sigma_f} = \frac{\text{cov}(a_i, \frac{1}{2} \hat{a}_s + \frac{1}{2} \hat{a}_d)}{\sqrt{\sigma_a^2 \text{var}(\frac{1}{2} \hat{a}_s + \frac{1}{2} \hat{a}_d)}}$$

Predykcja wartości hodowlanych na podstawie wydajności krewnych

- matka i ojciec osobnika i są niespokrewnione

$$\text{cov}\left(a_i, \frac{1}{2}\widehat{a}_s + \frac{1}{2}\widehat{a}_d\right) = \text{cov}\left(a_i, \frac{1}{2}\widehat{a}_s\right) + \text{cov}\left(a_i, \frac{1}{2}\widehat{a}_d\right)$$

$$= \text{cov}\left(\frac{1}{2}a_s + \frac{1}{2}a_d, \frac{1}{2}\widehat{a}_s\right) + \text{cov}\left(\frac{1}{2}a_s + \frac{1}{2}a_d, \frac{1}{2}\widehat{a}_d\right)$$

$$= \text{cov}\left(\frac{1}{2}a_s, \frac{1}{2}\widehat{a}_s\right) + \text{cov}\left(\frac{1}{2}a_d, \frac{1}{2}\widehat{a}_d\right)$$

$$\begin{aligned} &= \frac{1}{4}\text{var}(\widehat{a}_s) + \frac{1}{4}\text{var}(\widehat{a}_d) = \frac{1}{4}\text{var}(b \cdot y_s) + \frac{1}{4}\text{var}(b \cdot y_d) = \frac{1}{4}\text{var}(h^2 \cdot y_s) + \frac{1}{4}\text{var}(h^2 \cdot y_d) = \frac{1}{4}h^4[\sigma_y^2 + \sigma_y^2] \\ &= \frac{1}{4}[r_s^2 + r_d^2]\sigma_a^2 \end{aligned}$$

$$r = \frac{\text{cov}(a, f)}{\sigma_a \sigma_f}$$

Predykcja wartości hodowlanych na podstawie wydajności krewnych

- matka i ojciec osobnika i są niespokrewnione

$$r = \frac{cov(a, f)}{\sigma_a \sigma_f}$$

$$\sigma_a \sigma_f = \sqrt{\sigma_a^2 var\left(\frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d\right)} = \sqrt{\sigma_a^2 \frac{1}{4}[r_s^2 + r_d^2]\sigma_a^2}$$

$$r = \frac{cov(a, f)}{\sigma_a \sigma_f} = \frac{\frac{1}{4}[r_s^2 + r_d^2]\sigma_a^2}{\sqrt{\sigma_a^2 \frac{1}{4}[r_s^2 + r_d^2]\sigma_a^2}} = \frac{1}{2}\sqrt{r_s^2 + r_d^2}$$

Predykcja wartości hodowlanych na podstawie wydajności krewnych

- Powtarzalność (repeatability)

$$r^2 = \frac{r_s^2 + r_d^2}{4}$$

- Oczekiwany postęp selekcyjny (expected response to selection)

$$R = i r^2 \sigma_a$$

Predykcja wartości hodowlanej cechy A na podstawie wydajności cechy B

- Wartość hodowlaną cechy A można przewidzieć na podstawie wartości hodowlanej cechy B skorelowanej z nią genetycznie
- Np. cechy pokroju i wydajności

	ST	SR	BD	DF	RA	TW	RL	FA	FU	UH	UW	UC	UD	TP	FS	2ND	54MO	84MO	THL	FHL	MY
ST ²		.73	.81	.34	-.11	.67	.02	.30	.32	.24	.28	.21	.32	.19	.66	.14	.09	-.01	.06	.06	.01
SR	.58		.92	-.10	-.17	.72	-.13	.41	.29	.21	.33	.19	.09	.20	.55	-.05	-.08	-.17	-.11	-.11	-.04
BD	.65	.77		.23	-.15	.71	-.03	.30	.26	.20	.31	.18	.05	.20	.60	.01	.00	-.16	-.06	-.09	.07
DF	.16	-.15	.08		-.05	.13	.31	-.15	.00	.16	.18	.19	-.03	.13	.37	.40	.46	.25	.40	.23	.52
RA	.02	-.04	-.03	.00		-.24	-.04	-.20	-.29	-.34	-.28	-.24	-.19	-.24	-.34	.05	.04	.11	.08	.04	.13
TW	.43	.46	.47	.03	-.08		.01	.27	.31	.31	.42	.16	.17	.18	.58	.04	.01	-.08	-.02	-.03	.02
RL	-.04	-.10	-.05	.14	-.01	-.03		-.43	-.08	-.10	-.11	-.11	-.04	-.06	-.04	.00	.00	.01	.00	-.02	.06
FA	.14	.20	.17	-.03	-.05	.14	-.21		.28	.23	.23	.33	.25	.23	.43	.05	.05	.06	.08	.11	-.06
FU	.14	.16	.13	.02	-.12	.14	-.05	.14		.64	.63	.51	.79	.63	.73	.24	.29	.26	.29	.42	-.23
UH	.14	.14	.16	.14	-.11	.16	-.07	.14	.41		.90	.42	.45	.37	.67	.25	.30	.27	.32	.33	.06
UW	.17	.20	.21	.15	-.07	.24	-.09	.15	.37	.71		.44	.32	.47	.71	.28	.31	.22	.31	.28	.16
UC	.08	.06	.07	.12	-.09	.08	.01	.08	.31	.28	.26		.48	.76	.65	.26	.28	.25	.30	.34	-.01
UD	.15	.01	-.03	-.06	-.08	.05	-.02	.07	.43	.17	.08	.31		.43	.59	.18	.18	.29	.24	.44	-.41
TP	.08	.07	.08	.07	-.08	.08	.00	.08	.40	.23	.23	.46	.31		.64	.25	.25	.19	.24	.28	-.03
FS	.42	.36	.41	.27	-.15	.33	-.09	.28	.50	.49	.50	.43	.31	.42		.37	.39	.30	.39	.41	.06
2ND	.04	.04	.05	.06	.00	.03	-.01	.03	.06	.08	.08	.07	.01	.06	.14		.89	.78	.87	.83	.33
54MO	.04	.02	.04	.08	.00	.03	-.01	.03	.08	.10	.10	.08	.04	.08	.16	.54		.91	.99	.89	.51
84MO	.02	.00	.00	.05	-.01	.01	-.01	.02	.08	.07	.06	.06	.06	.06	.10	.19	.36		.95	.95	.28
THL	.04	.02	.03	.09	-.01	.02	-.01	.04	.10	.11	.11	.10	.06	.09	.19	.56	.77	.72		.93	.44
FHL	.03	.01	.01	.04	-.02	.01	-.02	.04	.12	.09	.08	.10	.10	.09	.16	.55	.75	.72	.98		.08
MY	.07	.04	.10	.28	.04	.06	.02	.00	-.07	.10	.16	.05	-.23	.03	.14	.12	.18	.09	.19	.03	

¹Genetic correlations above diagonal.

²ST = Stature, SR = strength, BD = depth of body, DF = dairy form, RA = rump angle, TW = thurl width, RL = rear leg set, FA = foot angle, FU = fore udder attachment, UH = udder height, UW = udder width, UC = udder cleft, UD = udder depth, TP = teat placement, FS = final score, 2ND = second calving, 54MO = 54 mo of age, 84MO = 84 mo of age, THL = true herd life, FHL = functional herd life, MY = milk yield.

Pretykcyja wartoŃci hodowlanej cechy A na podstawie wydajnoŃci cechy B

- Źródło informacji – pojedyncza wydajnoŃć os. i

$$\hat{a}_i = b(y_i - \beta_i)$$

- Źródło informacji – wydajnoŃci córek osobnika i

$$\hat{a}_i = b(\bar{y}_i - \beta_i)$$

- Źródło informacji – wartoŃci hodowlane rodziców osobnika i

$$\hat{a}_i = \frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d$$

- Źródło informacji – fenotyp innej cechy osobnika i

$$\hat{a}_{ix} = b(y_i - \beta_i)$$

Predykcja wartości hodowlanej cechy A na podstawie wydajności cechy B

- b współczynnik regresji prawdziwej wartości hodowlanej (a) na wartość fenotypową (y)

$$b = \frac{\text{cov}(a, y)}{\text{var}(y)} = \frac{\text{cov}(a, a + e)}{\text{var}(y)} = \frac{\sigma_a^2}{\sigma_y^2} = h^2$$

$$b = \frac{\text{cov}(a_x, y)}{\text{var}(y)}$$

$$b = \frac{r_a(x, y) \sigma_a(x) \sigma_a(y)}{\sigma_y^2} = r_a(x, y) h_x h_y \frac{\sigma_x}{\sigma_y}$$

zależny od:
korelacji genetycznej
odziedziczalności
wariancji fenotypowych

- $r_a(x, y)$ korelacja genetyczna pomiędzy x i y

Predykcja wartości hodowlanej cechy A na podstawie wydajności cechy B

- Dokładność (accuracy) predykcji = korelacja pomiędzy kryterium selekcyjnym (y), a prawdziwą wartością hodowlaną (a)

$$r = \frac{\text{cov}(\mathbf{a}, \mathbf{y})}{\sigma_a \sigma_y} = \frac{\sigma_a^2}{\sigma_a \sigma_y} = \frac{\sigma_a}{\sigma_y} = h$$

$$r = \frac{\text{cov}(\mathbf{a}_x, \mathbf{y})}{\sigma_a(x) \sigma_y} = \frac{r_a(x, y) \sigma_a(x) \sigma_a(y)}{\sigma_a(x) \sigma_y} = r_a(x, y) h_y$$

zależna od:
korelacji genetycznej między cechami
odziedziczalności cechy mierzonej

Indeks selekcyjny

- Selection index
- W praktyce dla osobnika dostępne są jednocześnie różne źródła informacji
- Indeks selekcyjny łączy wszystkie źródła informacji dostępne do predykcji wartości hodowlanej osobnika
- Indeks selekcyjny pozwala na uzyskanie najbardziej dokładnej predykcji

Indeks selekcyjny

$$I_i = \hat{a}_i = b_1(y_1 - \beta_1) + b_2(y_2 - \beta_2) + b_3(y_3 - \beta_3)$$

- I_i indeks selekcyjny osobnika $i \rightarrow$ model regresji liniowej dla różnych źródeł informacji
- y_i różne źródła informacji (np. wydajność własna + wydajności rodziców)
- b_i wagi = współczynniki regresji dla poszczególnych źródeł informacji
- Zastosowanie indeksu selekcyjnego do predykcji wartości hodowlanej polega na obliczeniu b_i

Indeks selekcyjny

$$I_i = \hat{a}_i = b_1(y_1 - \beta_1) + b_2(y_2 - \beta_2) + b_3(y_3 - \beta_3)$$

- Indeks selekcyjny = estymator a_i
- Własności indeksu selekcyjnego
 - Minimalizuje **średni błąd predykcji** $(\hat{a}_i - a_i)^2$
 - Maksymalizuje $r(\hat{a}_i, a_i)$
 - Maksymalizuje prawdopodobieństwo uzyskania poprawnego = prawdziwego rankingu ocenianych osobników

Indeks selekcyjny

- Dokładność (accuracy) predykcji = korelacja pomiędzy kryterium selekcyjnym (I), a prawdziwą wartością hodowlaną (a)

$$r = \frac{cov(\mathbf{a}, \mathbf{y})}{\sigma_a \sigma_y} = \frac{\sigma_a^2}{\sigma_a \sigma_y} = \frac{\sigma_a}{\sigma_y} = h$$

$$r = \frac{cov(\mathbf{a}, \mathbf{I})}{\sigma_a \sigma_I}$$

Indeks selekcyjny

$$r = \frac{cov(\mathbf{a}, \mathbf{I})}{\sigma_a \sigma_I}$$

$$cov(a_i, I) = \sum_{j=1}^M b_j \sigma_{a_{ij}} = \sigma_I^2$$

$$r = \frac{\sigma_I^2}{\sigma_a \sigma_I} = \sqrt{\frac{\sigma_I^2}{\sigma_a^2}} = \sqrt{\frac{\sum_{j=1}^M b_j \sigma_{a_{ij}}}{\sigma_a^2}}$$

- M liczba źródeł informacji w indeksie selekcyjnym

Dziękuję za uwagę