

METODY STATYSTYCZNE W BIOLOGII

1. Wykład wstępny
2. Populacje i próby danych
3. Testowanie hipotez i estymacja parametrów
4. Planowanie eksperymentów biologicznych
5. Najczęściej wykorzystywane testy statystyczne
6. Podsumowanie materiału, wspólna analiza przykładów, dyskusja
7. Regresja liniowa
8. Regresja nieliniowa
9. Określenie jakości dopasowania równania regresji liniowej i nieliniowej
10. Korelacja
11. Elementy statystycznego modelowania danych - EDA
12. Porównywanie modeli
- 13. Analiza wariancji**
14. Analiza kowariancji
15. Podsumowanie materiału, wspólna analiza przykładów, dyskusja

ANALIZA WARIANCJI

- 1. Model jednoczynnikowy**
- 2. Model dwuczynnikowy**
- 3. Model hierarchiczny**

EGZAMIN

ANALIZA WARIANCJI – ZAŁOŻENIA

1. Zmienna zależna (y) - rozkład normalny
2. Jednakowa / proporcjonalna liczebność grup

MODEL JEDNOCZYNNIKOWY

1. Zawartość azotu w trzcinie (% suchej masy)
2. 3 lokalizacje (A, B, C), pomiar w 1996 r.
3. Flowermere, hrabstwo Cambridge

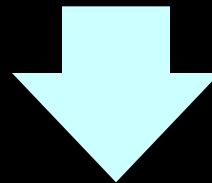
A	B	C
3.06	3.41	2.92
2.60	3.23	2.88
2.55	3.93	3.25
2.42	3.74	2.64
2.35	3.18	3.28



MODEL JEDNOCZYNNIKOWY – PRÓBA DANYCH

Obserwowana zmienność cechy

3.06	3.41	2.92
2.60	3.23	2.88
2.55	3.93	3.25
2.42	3.74	2.64
2.35	3.18	3.28



Różne lokalizacje

3.06	3.41	2.92
2.60	3.23	2.88
2.55	3.93	3.25
2.42	3.74	2.64
2.35	3.18	3.28

MODEL JEDNOCZYNNIKOWY – MODEL

MODEL ANALIZY WARIANCJI:

$$\text{Zawartość N} = \mu + \text{lokalizacja} + e$$

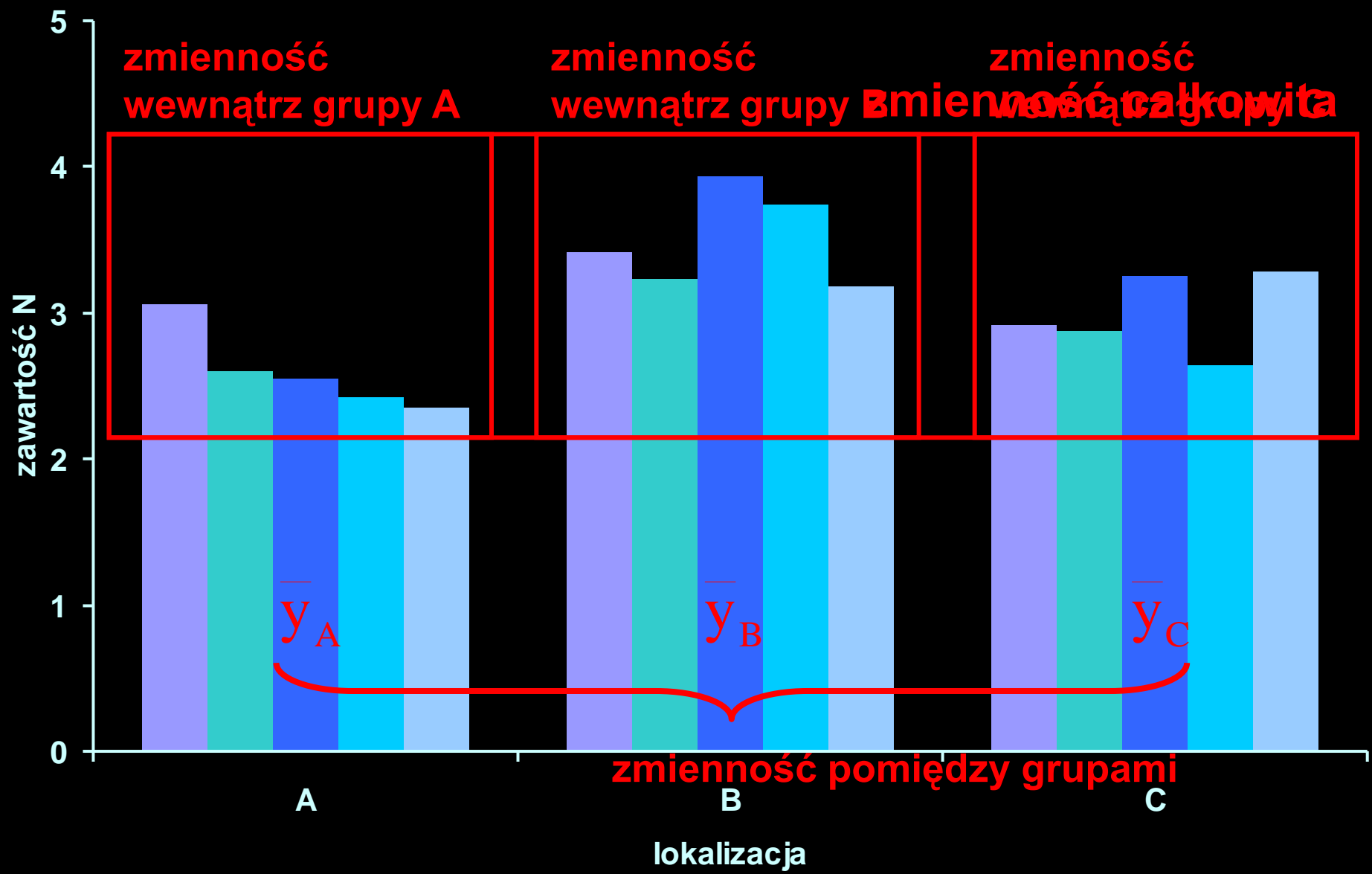
The diagram shows the equation $\text{Zawartość N} = \mu + \text{lokalizacja} + e$ with three red circles highlighting the terms μ , lokalizacja , and e . Red arrows point from these circles to three separate boxes below. The first box points to μ , the second to e , and the third to lokalizacja .

Efekty wspólne dla wszystkich obserwacji
Nie wpływają na zmienność

Efekty niezmierzone w doświadczeniu
Wpływają na zmienność

Analizowany czynnik
Testujemy czy wpływa na zmienność ???

MODEL JEDNOCZYNNIKOWY – MODEL



MODEL JEDNOCZYNNIKOWY – MODEL

SCHEMAT JEDNOCZYNNIKOWEJ ANALIZY WARIANCJI

$$\sigma_{\text{azot}}^2 = \sigma_{\text{lokalizacja}}^2 + \sigma_e^2$$

ŹRÓDŁO ZMIENNOŚCI	SUMA KWADRATÓW	STOPNIE SWOBODY	ŚREDNIA KWADRAT
Pomiędzy gr. (lokalizacja)	$\sum_{i=1}^{N_g} n_i (\bar{y}_i - \bar{y})^2$	$N_g - 1$	$\frac{\sum_{i=1}^{N_g} n_i (\bar{y}_i - \bar{y})^2}{N_g - 1}$
Wewnątrz gr. błąd	$\sum_{i=1}^{N_g} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$N - N_g$	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{N - N_g}$
Całkowita	$\sum_{i=1}^N (y_i - \bar{y})^2$	$N - 1$	$\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N - 1}$

MODEL JEDNOCZYNNIKOWY – TEST F

- **Testowane hipotezy**

H_0 : lokalizacje nie wpływają na zawartość azotu

H_1 : lokalizacje wpływają na zawartość azotu

$$H_0: \sigma_{\text{lok}}^2 \leq \sigma_e^2$$

$$H_1: \sigma_{\text{lok}}^2 > \sigma_e^2$$

- **Test F:**

$$F = \frac{\sum_{i=1}^{N_g} n_i (\bar{y}_i - \bar{y})^2}{N_g - 1} \bigg/ \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{N - N_g}$$

średnia zmienność wartości y spowodowana różnymi lokalizacjami

średnia zmienność wartości y nie wyjaśniona przez różne lokalizacje = średni błąd

3. Wybór i obliczenie wartości testu statystycznego

$$F = \frac{\sum_{i=1}^{N_g} n_i (\bar{y}_i - \bar{y})^2}{N_g - 1} = \frac{2.04}{3 - 1} = 11.95$$
$$F = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{N - N_g} = \frac{1.03}{15 - 3}$$

4. Określenie rozkładu testu: $\sim F_{N_g - 1, N - N_g}$

5. Obliczenie wartości α_t : $\alpha_t = 0.00139$

6. Decyzja: $\alpha_t < \alpha_{\max}$

~~H_0~~

H_1

lokalizacje wpływają na zawartość azotu w suchej masie trzciny

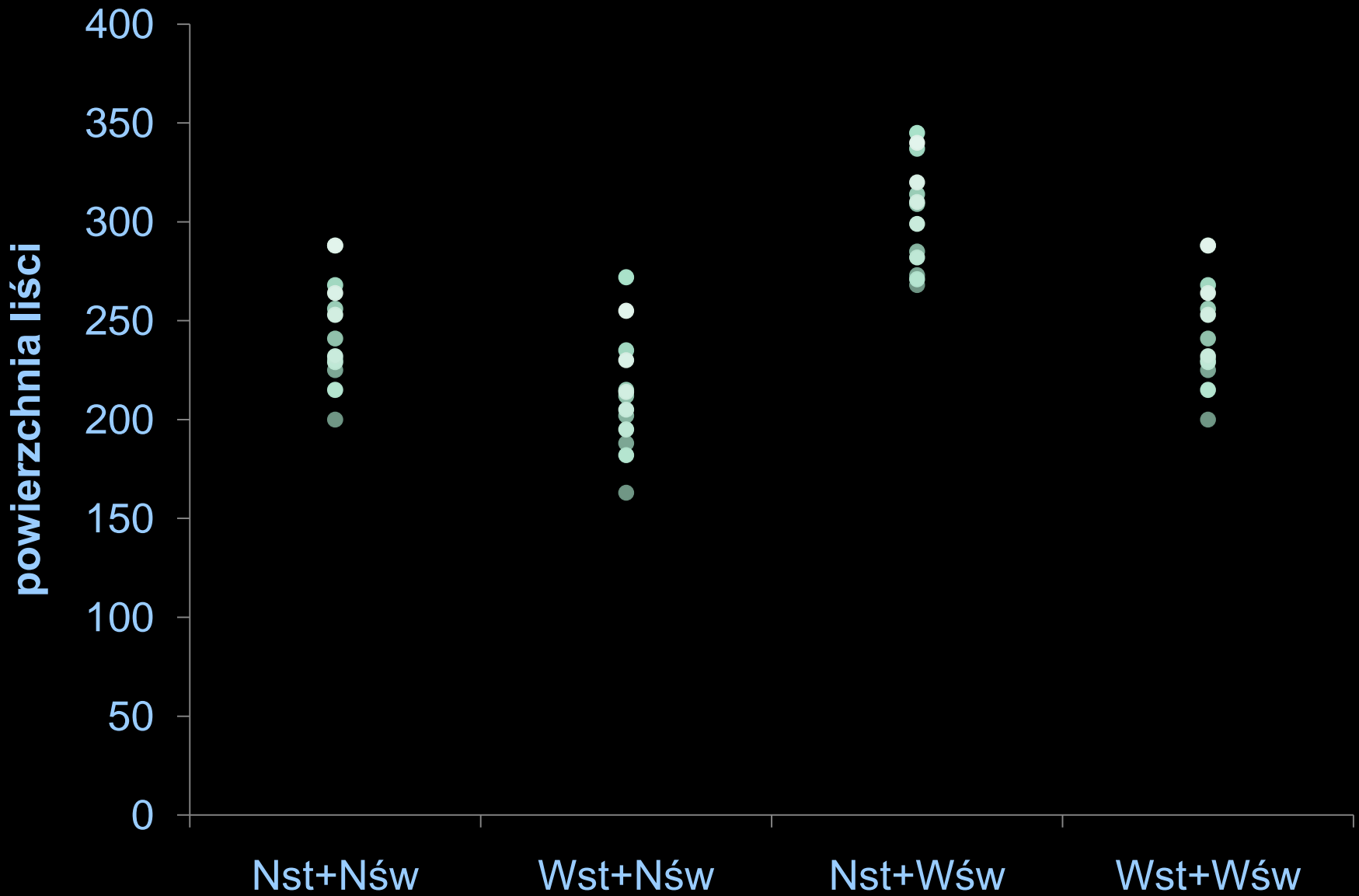
MODEL DWUCZYNNIKOWY

1. Wzrost soi - powierzchnia liści
2. Stres mechaniczny
3. Nasłonecznienie



	niski poziom stresu	wysoki poziom stresu
słabe nasłonecznienie	200, 225, 230, ... , 264, 288	163, 188, 202, ... , 230, 255
dobrze nasłonecznienie	268, 273, 285, ... , 320, 349	200, 225, 230, ... , 264, 288

MODEL DWUCZYNNIKOWY – PRÓBA DANYCH



MODEL ANALIZY WARIANCJI:

- bez interakcji

powierzchnia liści = μ + stres + słońce + e

$$\sigma_{powierzchnia}^2 = \sigma_{stres}^2 + \sigma_{slonce}^2 + \sigma_e^2$$

$$H_0 : \quad \sigma_{stres}^2 \leq \sigma_e^2 \quad \sigma_{slonce}^2 \leq \sigma_e^2$$

$$H_1 : \quad \sigma_{stres}^2 > \sigma_e^2 \quad \sigma_{slonce}^2 > \sigma_e^2$$

- z interakcją

powierzchnia liści = μ + stres + słońce + stres*słońce + e

$$\sigma_{powierzchnia}^2 = \sigma_{stres}^2 + \sigma_{slonce}^2 + \sigma_{stres*slonce}^2 + \sigma_e^2$$

$$H_0 : \quad \sigma_{stres}^2 \leq \sigma_e^2 \quad \sigma_{slonce}^2 \leq \sigma_e^2 \quad \sigma_{stres*slonce}^2 \leq \sigma_e^2$$

$$H_1 : \quad \sigma_{stres}^2 > \sigma_e^2 \quad \sigma_{slonce}^2 > \sigma_e^2 \quad \sigma_{stres*slonce}^2 > \sigma_e^2$$

MODEL DWUCZYNNIKOWY – MODEL

ŹRÓDŁO ZMIENNOŚCI	SUMA KWADRATÓW	STOPNIE SWOBODY	ŚREDNIA KWADRAT
Pomiędzy gr. (stres)	$\sum_{i=1}^{N_{st}} n_i (\bar{y}_i - \bar{y})^2$	$N_{st} - 1$	$\frac{\sum_{i=1}^{N_{st}} n_i (\bar{y}_i - \bar{y})^2}{N_{st} - 1}$
Pomiędzy gr. (słońce)	$\sum_{j=1}^{N_{św}} n_j (\bar{y}_j - \bar{y})^2$	$N_{św} - 1$	$\frac{\sum_{j=1}^{N_{św}} n_j (\bar{y}_j - \bar{y})^2}{N_{św} - 1}$
Interakcja	$\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} n_{ij} (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$	$(N_{st} - 1)(N_{św} - 1)$	$\frac{\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} \sum_{k=1}^{n_{ij}} n_{ij} (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2}{N - N_{st} N_{św}}$
Wewnątrz gr. błąd	$\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2$	$N - N_{st} N_{św}$	$\frac{\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{N - N_{st} N_{św}}$
Całkowita	$\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y})^2$	$N - 1$	$\frac{\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y})^2}{N - 1}$

MODEL DWUCZYNNIKOWY – MODEL

$$H_0 : \sigma_{stres}^2 \leq \sigma_e^2 \quad H_1 : \sigma_{stres}^2 > \sigma_e^2$$

Pomiędzy gr.
(stres)

$$\sum_{i=1}^{N_{st}} n_i (\bar{y}_i - \bar{y})^2$$

$$N_{st} - 1$$

$$\frac{\sum_{i=1}^{N_{st}} n_i (\bar{y}_i - \bar{y})^2}{N_{st} - 1}$$

Pomiędzy gr.
(słońce)

F

Interakcja

Wewnątrz gr.
błąd

$$\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2$$

$$N - N_{st} N_{św}$$

$$\frac{\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{św}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{N - N_{st} N_{św}}$$

Całkowita

MODEL DWUCZYNNIKOWY – MODEL

$$H_0 : \sigma_{słońce}^2 \leq \sigma_e^2 \quad H_1 : \sigma_{słońce}^2 > \sigma_e^2$$

Pomiędzy gr.
(stres)

Pomiędzy gr.
(słońce)

$$\sum_{j=1}^{N_{\acute{s}w}} n_j (\bar{y}_j - \bar{y})^2$$

$$N_{\acute{s}w} - 1$$

$$\frac{\sum_{j=1}^{N_{\acute{s}w}} n_j (\bar{y}_j - \bar{y})^2}{N_{\acute{s}w} - 1}$$

Interakcja

Wewnątrz gr.
błąd

$$\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{\acute{s}w}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2$$

$$N - N_{st} N_{\acute{s}w}$$

$$\frac{\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{\acute{s}w}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{N - N_{st} N_{\acute{s}w}}$$

F

Całkowita

MODEL DWUCZYNNIKOWY – MODEL

$$H_0 : \sigma_{stres*slo\ncce}^2 \leq \sigma_e^2 \quad H_1 : \sigma_{stres*slo\ncce}^2 > \sigma_e^2$$

Pomiędzy gr.
(stres)

Pomiędzy gr.
(słońce)

Interakcja

Wewnątrz gr.
błąd

Całkowita

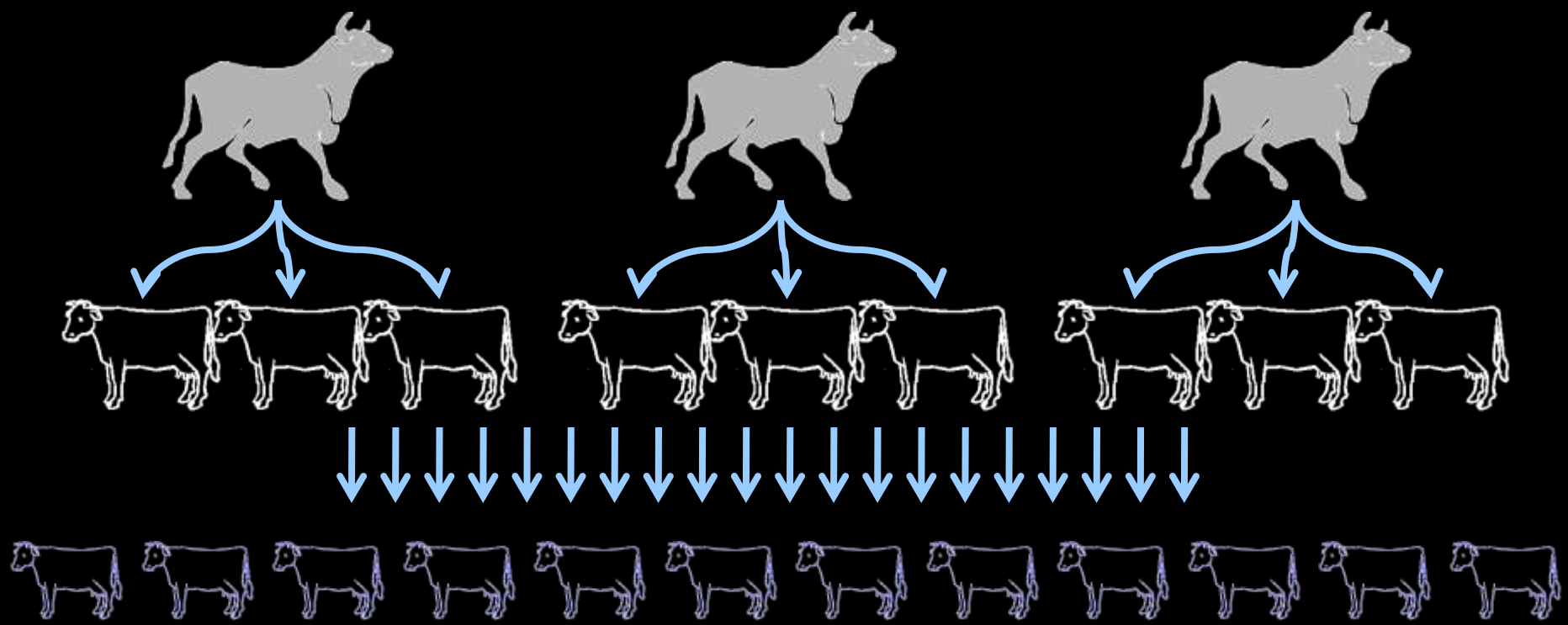
$$\begin{array}{l}
 \sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{s\acute{w}}} n_{ij} (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2 \quad (N_{st} - 1)(N_{s\acute{w}} - 1) \\
 \sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{s\acute{w}}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2 \quad N - N_{st}N_{s\acute{w}}
 \end{array}
 \begin{array}{l}
 \xrightarrow{F} \frac{\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{s\acute{w}}} \sum_{k=1}^{n_{ij}} n_{ij} (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2}{N - N_{st}N_{s\acute{w}}} \\
 \xrightarrow{F} \frac{\sum_{i=1}^{N_{st}} \sum_{j=1}^{N_{s\acute{w}}} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{N - N_{st}N_{s\acute{w}}}
 \end{array}$$

MODEL HIERARCHICZNY

MODEL HIERARCHICZNY – PRÓBA DANYCH

PRÓBA DANYCH

- 1. Liczba komórek somatycznych w mleku (SCS) – zdrowotność wymienia
- 2. Ojciec (genotyp)
- 3. Matka (genotyp)



MODEL ANALIZY WARIANCJI:

$$\text{SCS} = \mu + \text{buhaj} + \text{buhaj}(\text{krowa}) + e$$

$$\sigma_{SCS}^2 = \sigma_{\text{buhaj}}^2 + \sigma_{\text{buhaj}(\text{krowa})}^2 + \sigma_e^2$$

$$H_0 : \quad \sigma_{\text{buhaj}}^2 \leq \sigma_e^2 \qquad \sigma_{\text{buhaj}(\text{krowa})}^2 \leq \sigma_e^2$$

$$H_1 : \quad \sigma_{\text{buhaj}}^2 > \sigma_e^2 \qquad \sigma_{\text{buhaj}(\text{krowa})}^2 > \sigma_e^2$$

MODEL HIERARCHICZNY – PRÓBA DANYCH

ŹRÓDŁO ZMIENNOŚCI	SUMA KWADRATÓW	STOPNIE SWOBODY	ŚREDNIA KWADRAT
Pomiędzy gr. (buhaj)	$\sum_{i=1}^{N_b} n_i (\bar{y}_i - \bar{y})^2$	$N_b - 1$	$\frac{\sum_{i=1}^{N_b} n_i (\bar{y}_i - \bar{y})^2}{N_b - 1}$
Pomiędzy krowami wewnątrz buhajów	$\sum_{i=1}^{N_b} \sum_{j=1}^{N_k} n_{ij} (\bar{y}_{ij} - \bar{y}_i)^2$	$N_b (N_k - 1)$	$\frac{\sum_{i=1}^{N_b} \sum_{j=1}^{N_k} n_{ij} (\bar{y}_{ij} - \bar{y}_i)^2}{N_b (N_k - 1)}$
Wewnątrz gr. błąd	$\sum_{i=1}^{N_b} \sum_{j=1}^{N_k} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2$	$N - N_b N_k$	$\frac{\sum_{i=1}^{N_b} \sum_{j=1}^{N_k} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{N - N_b N_k}$
Całkowita	$\sum_{i=1}^{N_b} \sum_{j=1}^{N_k} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y})^2$	$N - 1$	$\frac{\sum_{i=1}^{N_b} \sum_{j=1}^{N_k} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y})^2}{N - 1}$

ANALIZA WARIANCJI

1. Model jednoczynnikowy

2. Model dwuczynnikowy

3. Model hierarchiczny

EGZAMIN