

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
12. Porównywanie modeli
13. Analiza wariancji
14. Analiza kowariancji
15. Podsumowanie materiału, wspólna analiza przykładów, dyskusja

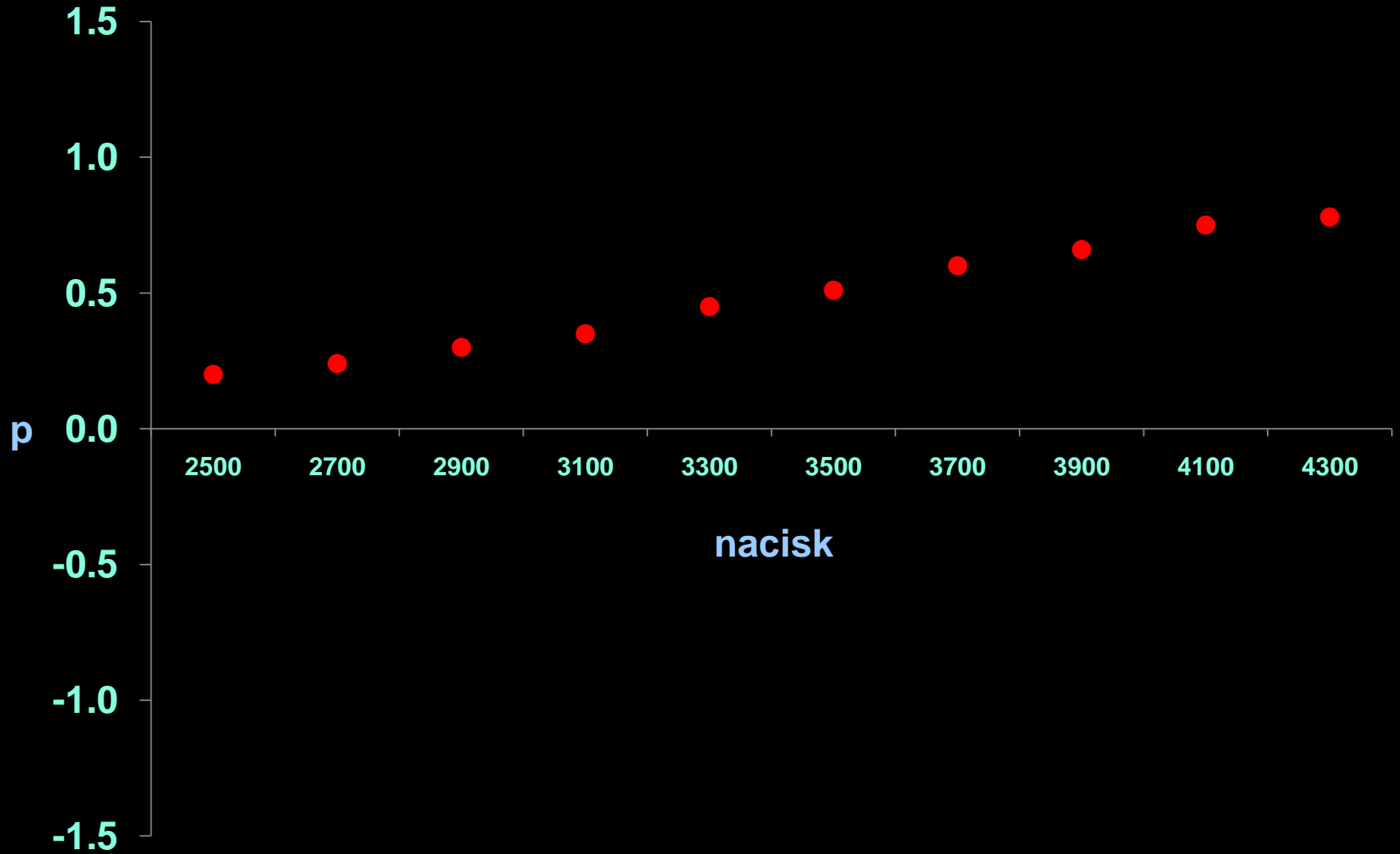
1. Dane pochodzące z rozkładu dwumianowego - przykład
2. Transformacje danych dwumianowych
 - logit
 - probit
 - log-log
3. Równanie regresji logistycznej
 - równanie
 - estymacja parametrów równania
 - przykład

DANE Z ROZKŁADU DWUMIANOWEGO

nacisk	ilość całkow.	ilość uszkod.
2500	50	10
2700	70	17
	...	
4300	65	51



- 1. Badanie wytrzymałości złącz w samolotach**
- 2. Zastosowano różne siły nacisku**





	chore	zdrowe
gr. badawcza	21	2
gr. kontrolna	19	13

1. Grupa badawcza - pomieszczenie wypełniane dymem papierosowym co godz. przez 12 godz. dziennie
2. Grupa kontrolna - identyczne pomieszczenie bez dymu
3. Sekcja osobników po 1 roku badań - nowotwór płuc



PRÓBA DANYCH 2.

choroba [y]	grupa [x]
1	 0=b.
1	0
0	0
...	...
1	 1=k.
1	1
0	1

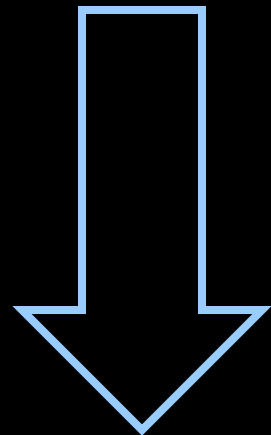
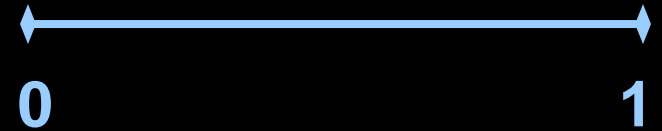
1. Dane nie pochodzą z rozkładu normalnego
2. Dane pochodzą z rozkładu dwumianowego
3. $Y \in \{ \text{chory, zdrowy} \}$
4. Prawdopodobieństwo zachorowania



TRANSFORMACJE DANYCH Z ROZKŁADU DWUMIANOWEGO

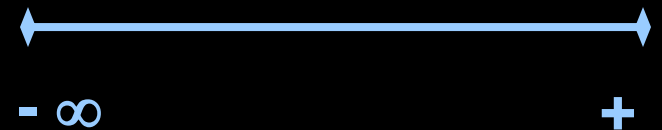
TRANSFORMACJA DANYCH

prawdopodobieństwo



transformacja

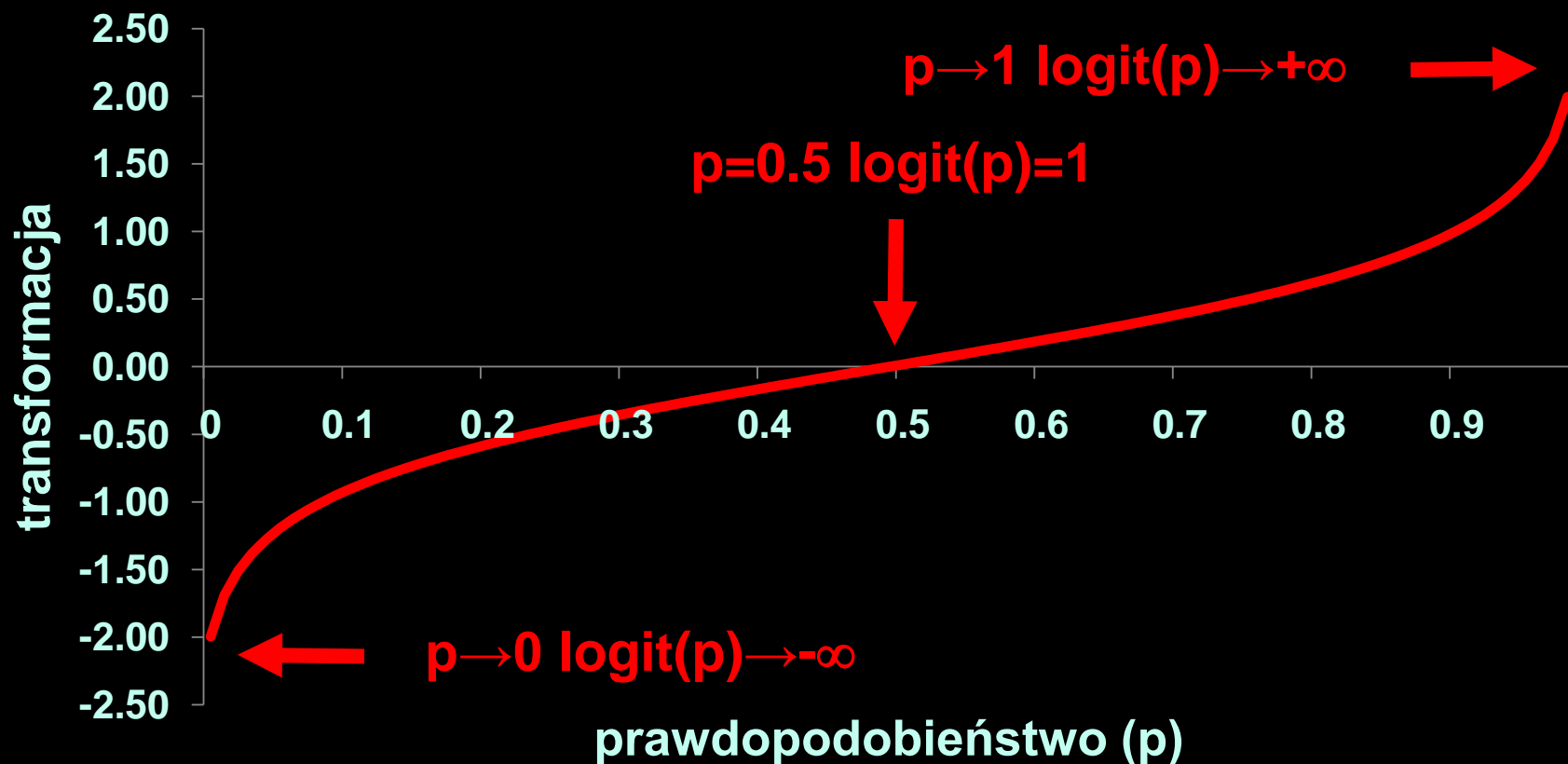
**funkcja
prawdopodobieństwa**



TRANSFORMACJA DANYCH

1. Transformacja logistyczna

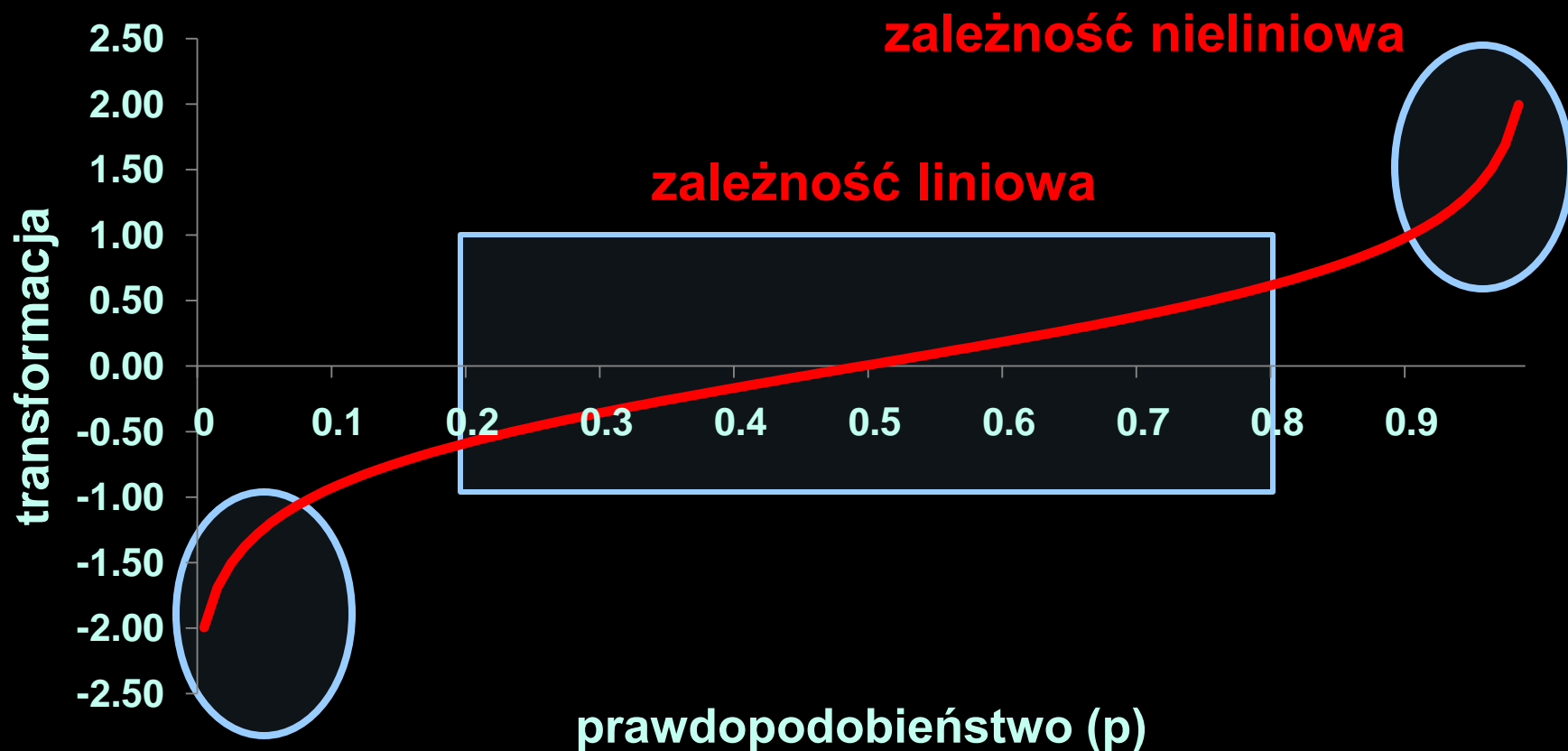
$$\text{logit}(p) = \log \frac{p}{1-p}$$



TRANSFORMACJA DANYCH

1. Transformacja logistyczna

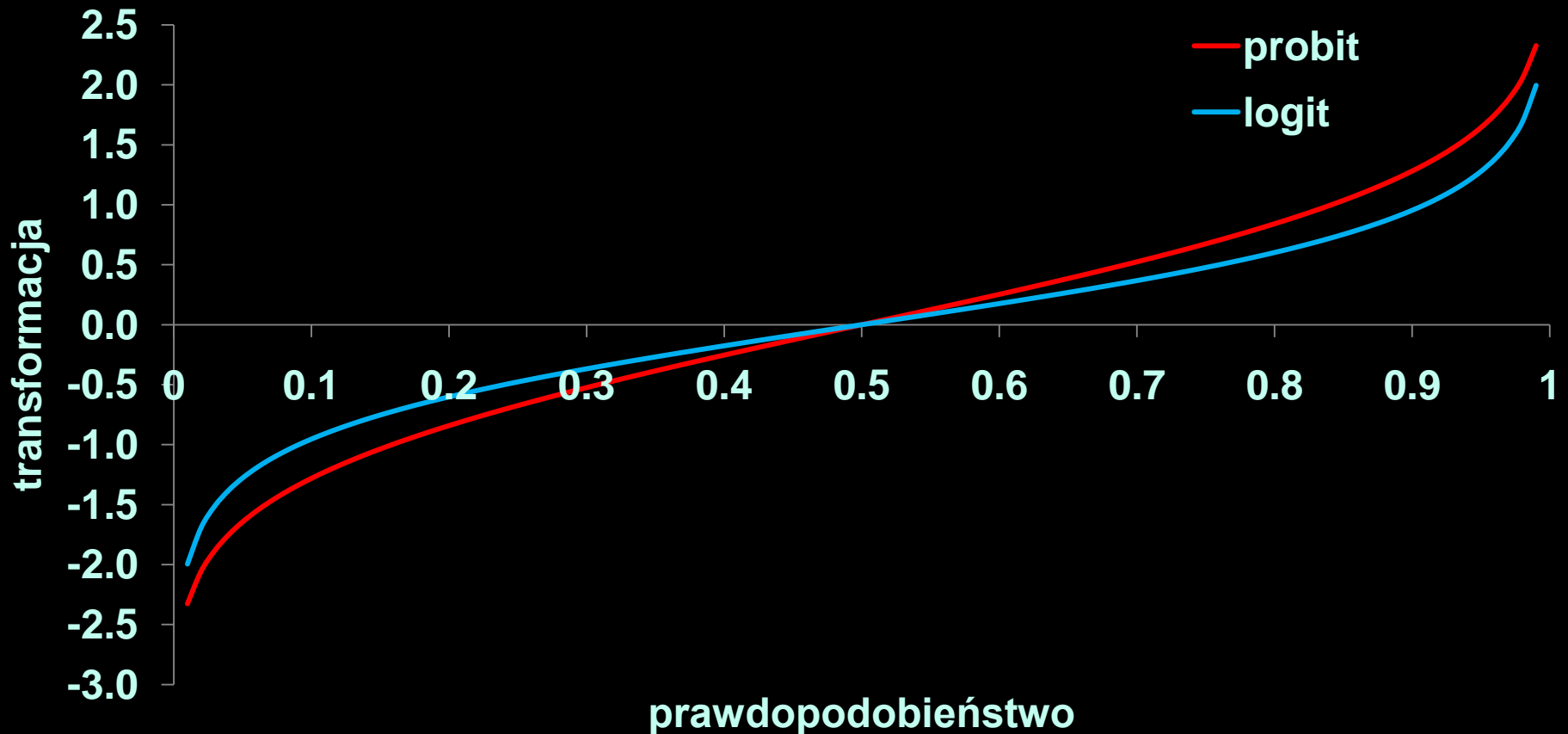
$$\text{logit}(p) = \log \frac{p}{1-p}$$



TRANSFORMACJA DANYCH

2. Transformacja probit

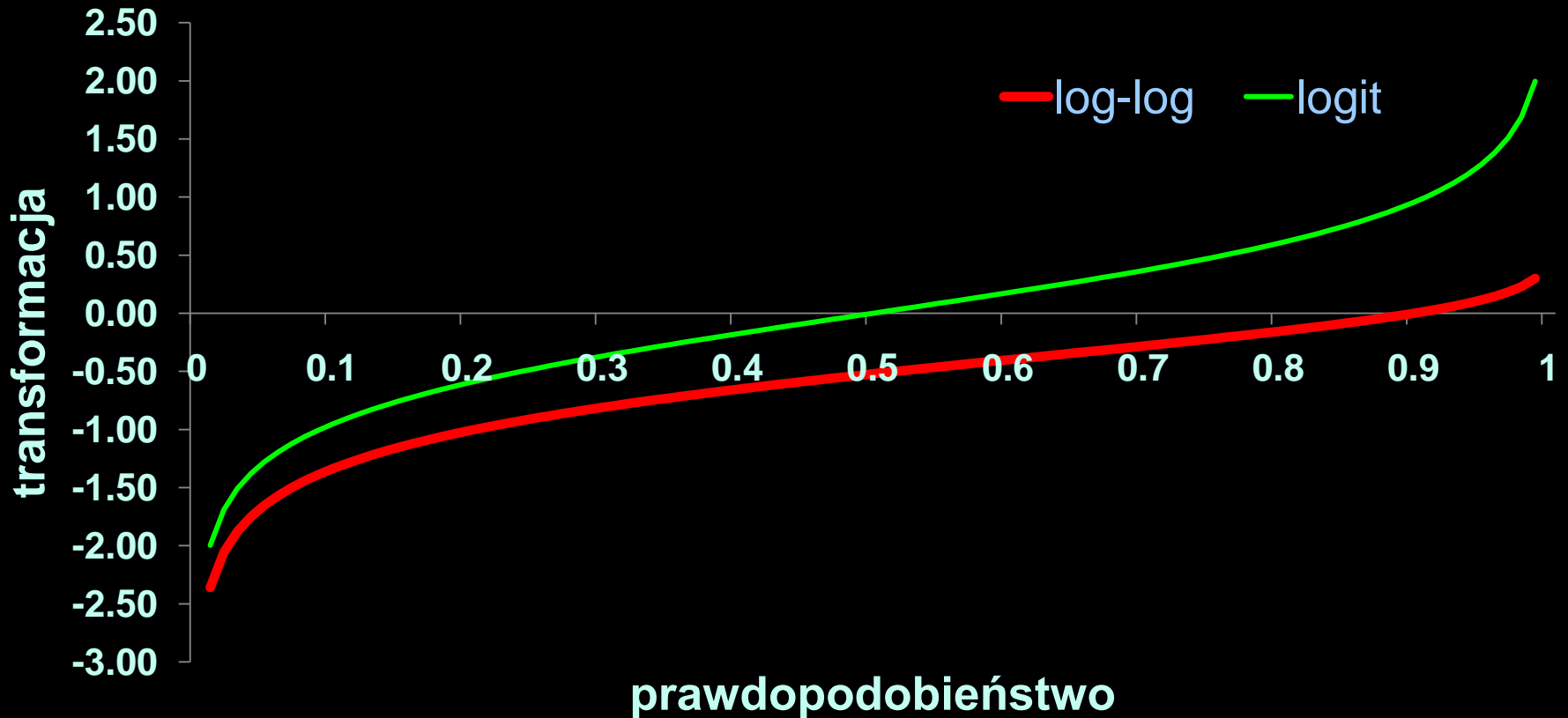
$$\text{probit}(p) = \Phi^{-1}(p)$$



TRANSFORMACJA DANYCH

3. Transformacja log-log

$$\log -\log(p) = \log[-\log(1-p)]$$



TRANSFORMACJA DANYCH

1. Transformacja logit - najczęściej stosowana
2. Transformacja probit - podobna do logit ale trudniejsza do obliczenia
3. Transformacja log-log - stosowana jeżeli modelowanie danych wymaga niesymetryczności względem $p=0.5$

RÓWNANIE REGRESJI LOGISTYCZNEJ

RÓWNANIE REGRESJI LOGISTYCZNEJ

$$\text{logit}(p) = \log \frac{p}{1-p} = \beta_0 + \beta_1 x$$

$$p = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

ESTYMACJA PARAMETRÓW

Iteracyjna metoda ważonych najmniejszych kwadratów:

$$\beta^{j+1} = \beta^j + \frac{\sum_{i=1}^N w_i (x_i - \bar{x})(y_i^j - \bar{y}^j)}{\sum_{i=1}^N w_i (x_i - \bar{x})^2}$$

$$w_i = n_i p_i (1 - p_i)$$

Fisher's method of scoring

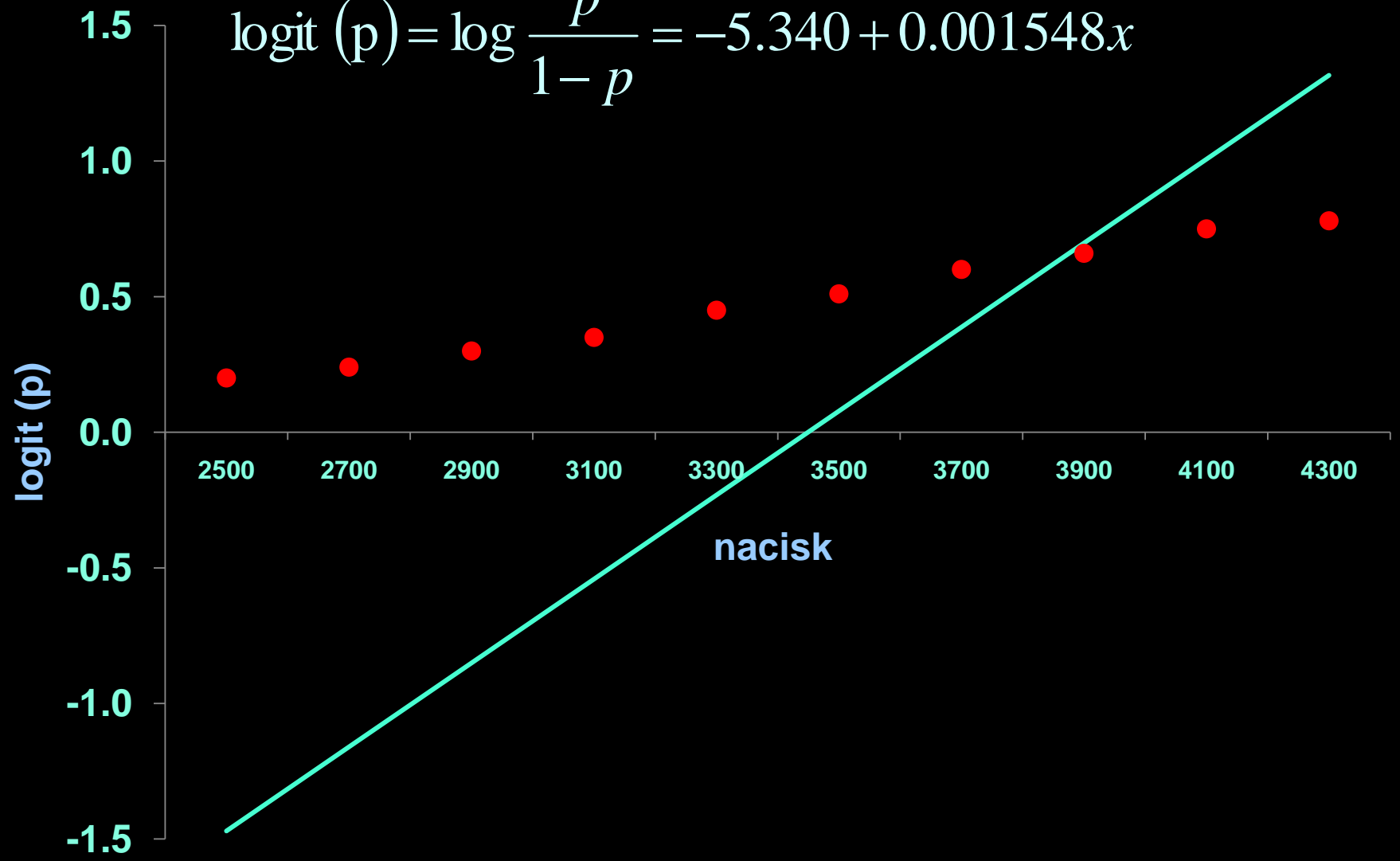
PRZYKŁAD

PRZYKŁAD

PRÓBA DANYCH 1.

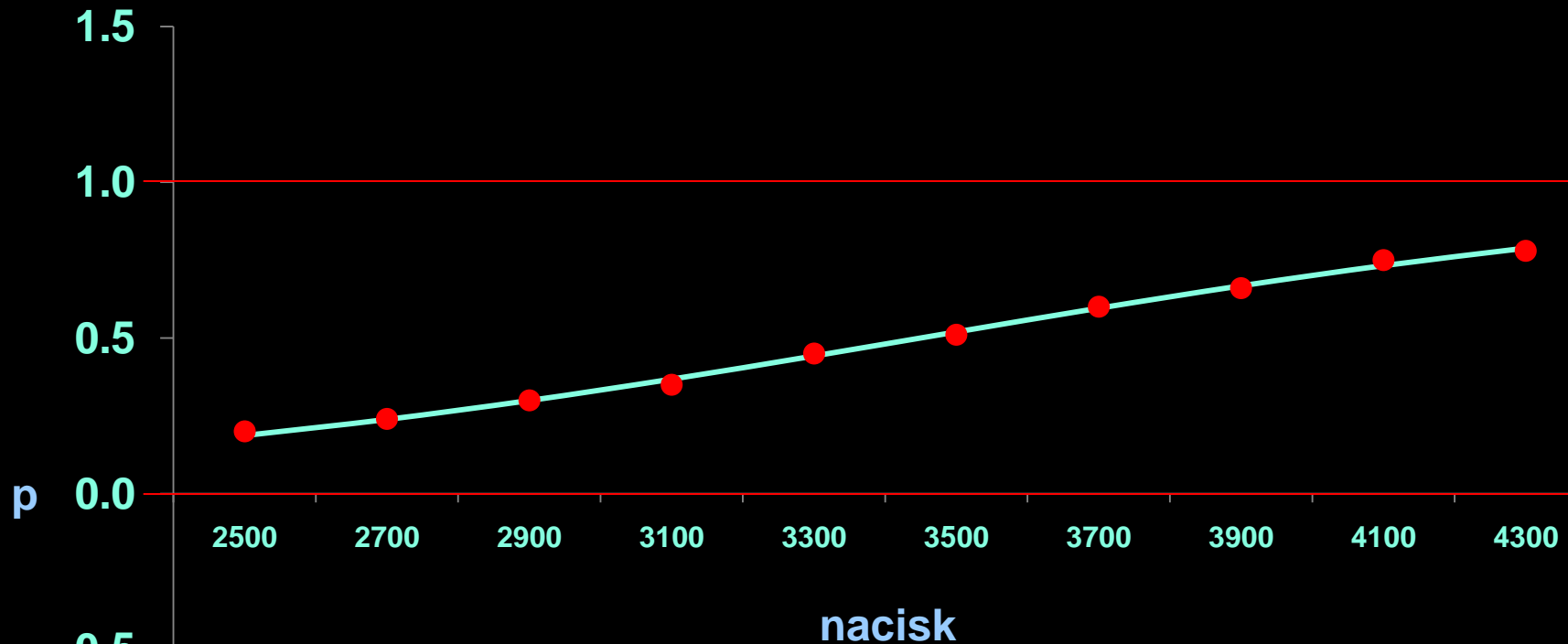
1. Równanie regresji

$$\text{logit}(p) = \log \frac{p}{1-p} = -5.340 + 0.001548x$$



PRZYKŁAD

PRÓBA DANYCH 1.



1. Prawdopodobieństwo uszkodzenia wzrasta wraz z naciskiem
2. Np. prawdopodobieństwo uszkodzenia dla nacisku 5000:

$$p = \frac{\exp(-5.340 + 0.001548 \cdot 5000)}{1 + \exp(-5.340 + 0.001548 \cdot 5000)} = 0.917$$

PRZYKŁAD

PRÓBA DANYCH 2.

choroba [y]	grupa [x]
1	0=b.
1	0
0	0
...	...
1	1=k.
1	1
0	1



1. Równanie regresji

$$\text{logit}(p) = \log \frac{p}{1-p} = 2.351 - 1.972x$$

2. Prawdopodobieństwo zachorowania jest mniejsze w g. kontrolnej

3. Prawdopodobieństwo zachorowania w grupie badawczej:

$$p_b = \frac{\exp(2.351 - 1.972 \cdot 0)}{1 + \exp(2.351 - 1.972 \cdot 0)} = 0.913$$

4. Prawdopodobieństwo zachorowania w grupie kontrolnej:

$$p_k = \frac{\exp(2.351 - 1.972 \cdot 1)}{1 + \exp(2.351 - 1.972 \cdot 1)} = 0.594$$

PRZYKŁAD

PRÓBA DANYCH 2.

choroba [y]	grupa [x]
1	0=b.
1	0
0	0
...	...
1	1=k.
1	1
0	1



1. Równanie regresji

$$\text{logit}(p) = \log \frac{p}{1-p} = 2.351 - 1.972x$$

2. Prawdopodobieństwo zachorowania jest mniejsze w g. kontrolnej

3. Prawdopodobieństwo zachorowania w grupie badawczej:

$$p_b = \frac{\exp(2.351 - 1.972 \cdot 0)}{1 + \exp(2.351 - 1.972 \cdot 0)} = 0.913$$

4. Prawdopodobieństwo zachorowania w grupie kontrolnej:

$$p_k = \frac{\exp(2.351 - 1.972 \cdot 1)}{1 + \exp(2.351 - 1.972 \cdot 1)} = 0.594$$

Open Access

Research article

Evidence for divergent selection between the molecular forms of *Anopheles gambiae*: role of predation

Abdoulaye Diabaté*^{1,2}, Roch K Dabiré², Kyle Heidenberger¹, Jacob Crawford¹, William O Lamp³, Lauren E Culler³ and Tovi Lehmann¹

Address: ¹Laboratory of Malaria and Vector Research, NIAID, National Institute of Health, MD 20852 USA, ²IRSS/Centre Muraz, Laboratoire de Parasitologie/Entomologie BP 39000, Université de Ouagadougou, Ouagadougou, Burkina Faso, ³Department of Entomology University of Maryland College Park, MD 20742-4454 USA

Email: Abdoulaye Diabaté* - a_diabate@hotmail.com; Roch K Dabiré - dabire_roch@univ-burkina.com



Table 2: Emergence success of the molecular forms in different habitats using a logistic regression accommodating cages variation

Source	Df	P	Estimate [ln(M/S)]
Intercept	1	0.84	-0.0232
Predator	1	<0.0001	+0.258
Habitat	1	0.019	+0.272a
Predator*Habitat	1	0.13	+0.08b
GroupCage (Habitat)	17	<0.0001	NA
Likelihood Ratio	11	<0.0001	NA

a: puddle over rice

b: predator over no predator in puddle/predator over no predator in rice

- 1. Dane pochodzące z rozkładu dwumianowego - przykład**
- 2. Transformacje danych dwumianowych**
 - **logit**
 - **probit**
 - **log-log**
- 3. Równanie regresji logistycznej**
 - **równanie**
 - **estymacja parametrów równania**
 - **przykład**