

# METODY STATYSTYCZNE W BIOLOGII

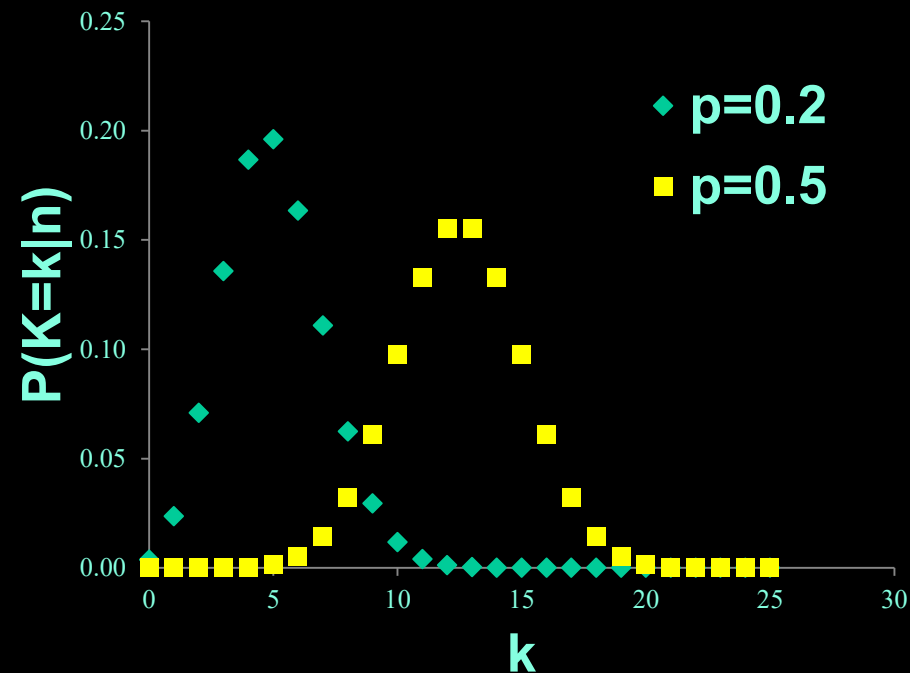
---

1. Wykład wstępny
2. Populacje i próby danych
3. Testowanie hipotez i estymacja parametrów
4. Planowanie eksperymentów biologicznych
5. Najczęściej wykorzystywane testy statystyczne I
6. Najczęściej wykorzystywane testy statystyczne II
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. Rozkład dwumianowy - przykład
2. Transformacje danych dwumianowych
  - logit
  - probit
  - log-log
3. Równanie regresji logistycznej
  - równanie
  - estymacja parametrów równania
  - przykład

# ROZKŁAD DWUMIANOWY

$$P(K = k | n, p) = \binom{n}{k} p^k (1-p)^{(n-k)}$$
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$



Przykłady ???

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



1. Badanie wytrzymałości złącz w samolotach
2. Zastosowano różne siły nacisku
3. Zmienna zależna – binarna → uszkodzenie / brak uszkodzenia
4. Zmienna niezależna – ciągła → nacisk





# Dane

	chore k	zdrowe n-k
gr. badawcza	21	2
gr. kontrolna	19	13

próba danych ②

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



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

1. Dane nie pochodzą z rozkładu normalnego
2. Dane pochodzą z rozkładu dwumianowego
3. Zmienna zależna – binarna  $\rightarrow$  chory / zdrowy
4. Zmienna niezależna – binarna  $\rightarrow$  dym / brak dymu



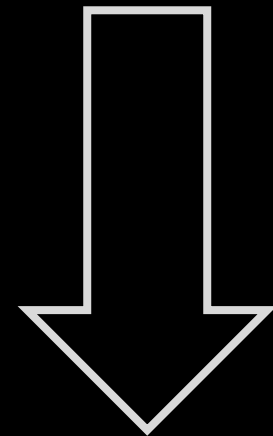
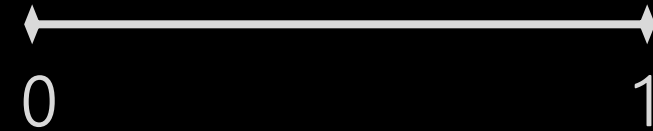


# TRANSFORMACJE DANYCH

# Transformacja danych

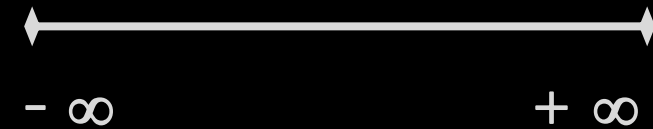
---

prawdopodobieństwo



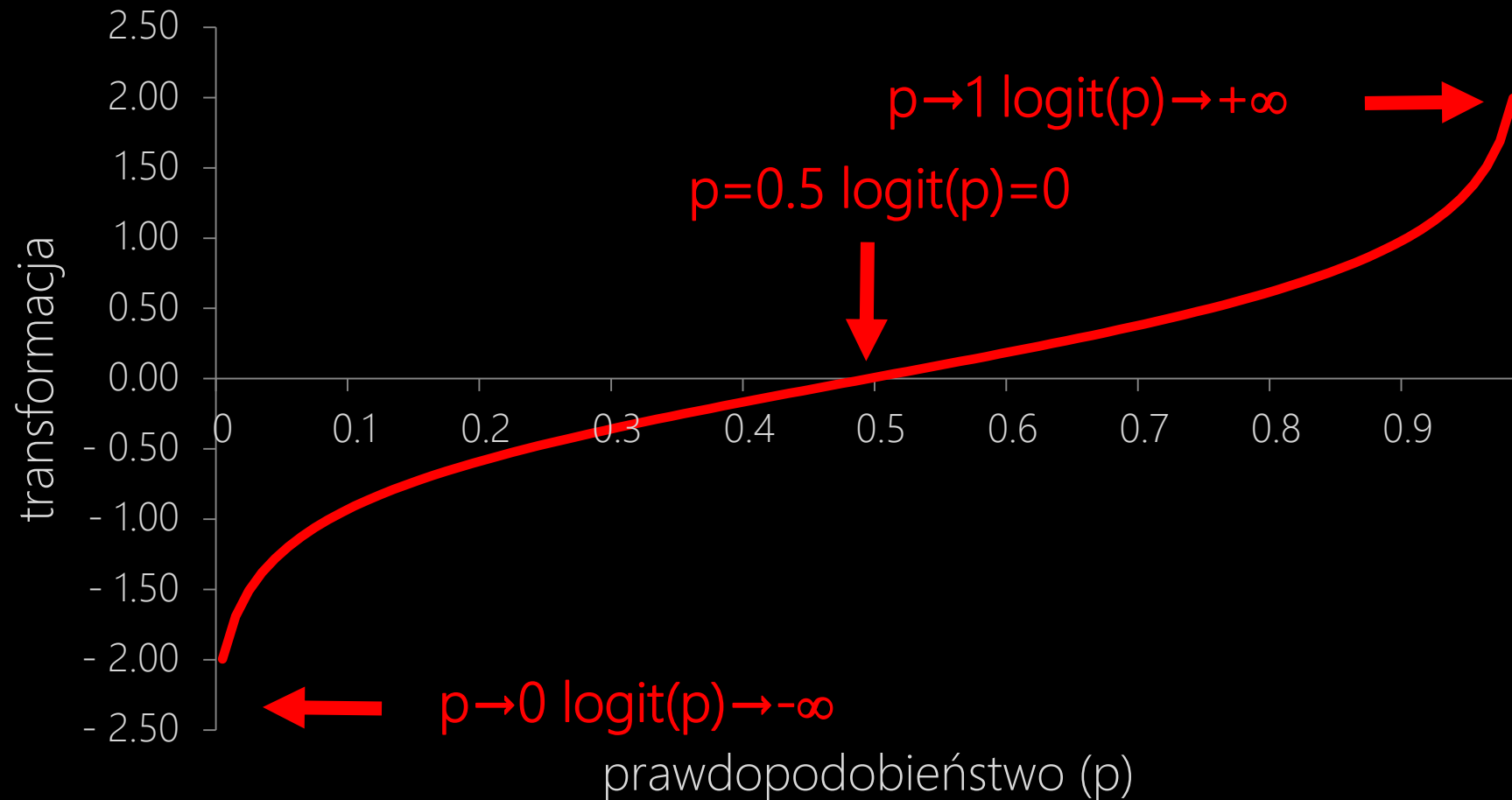
transformacja

funkcja  
prawdopodobieństwa



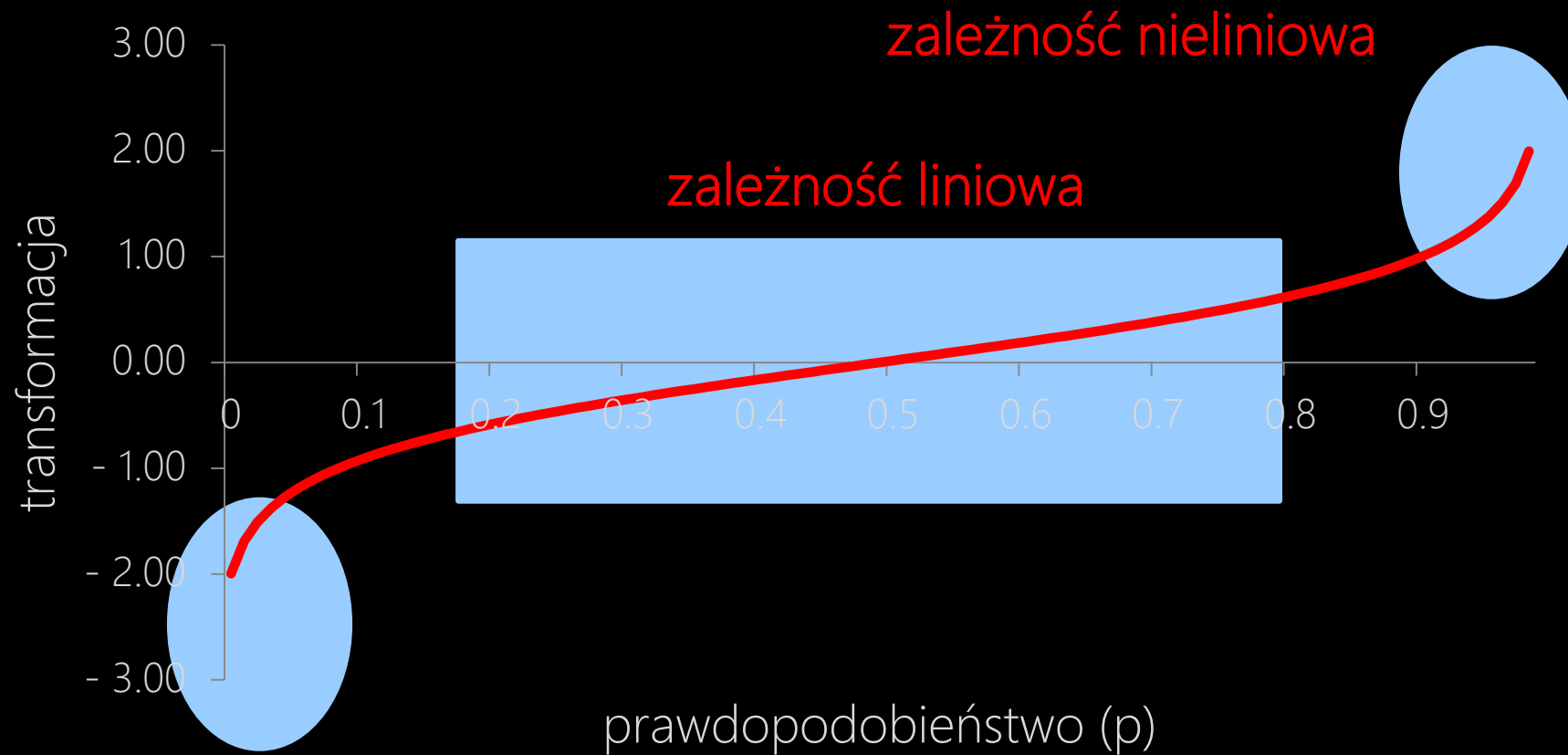
# Transformacja danych

1. Transformacja logistyczna:  $\text{logit}(p) = \ln \frac{p}{1-p}$



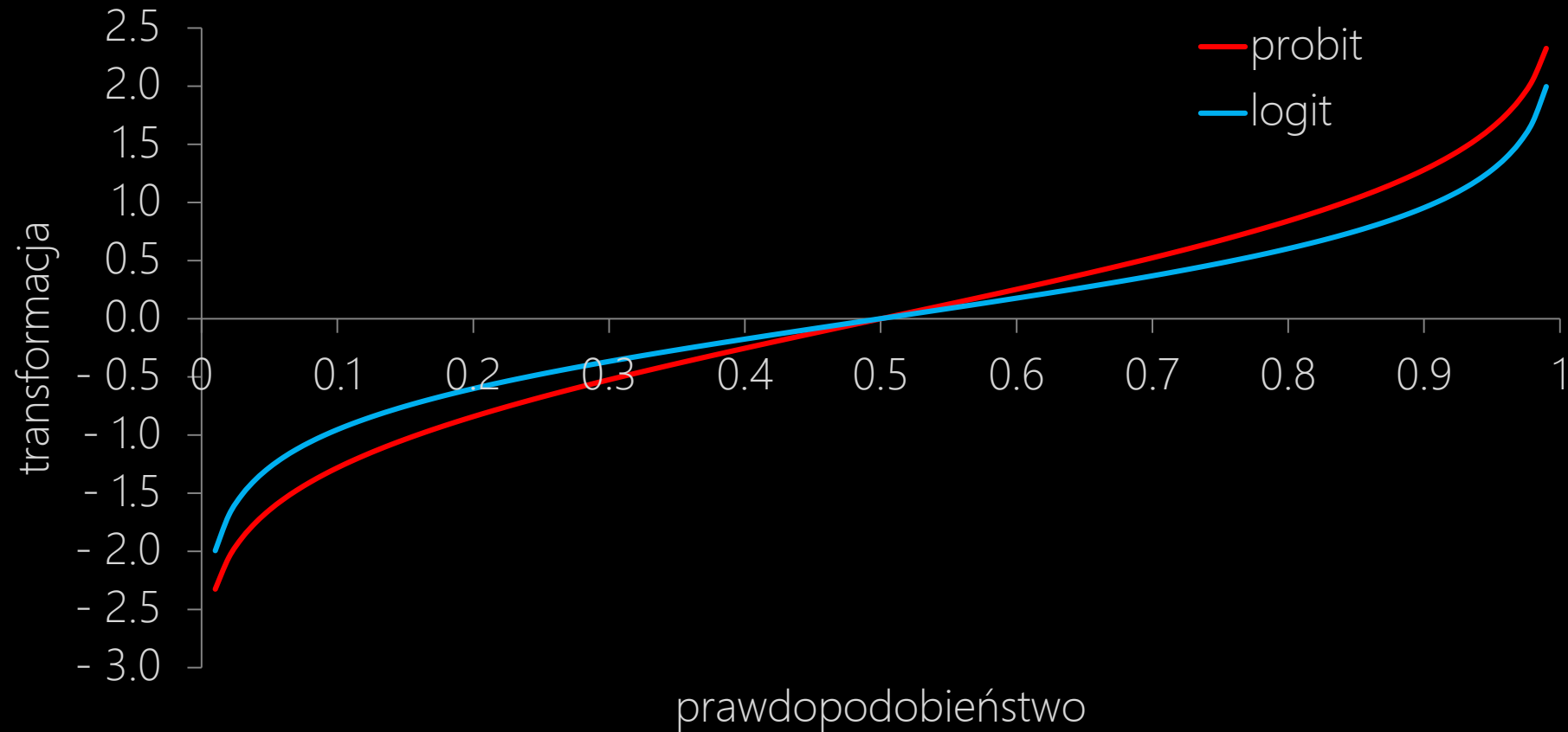
# Transformacja danych

1. Transformacja logistyczna:  $\text{logit}(p) = \ln \frac{p}{1-p}$



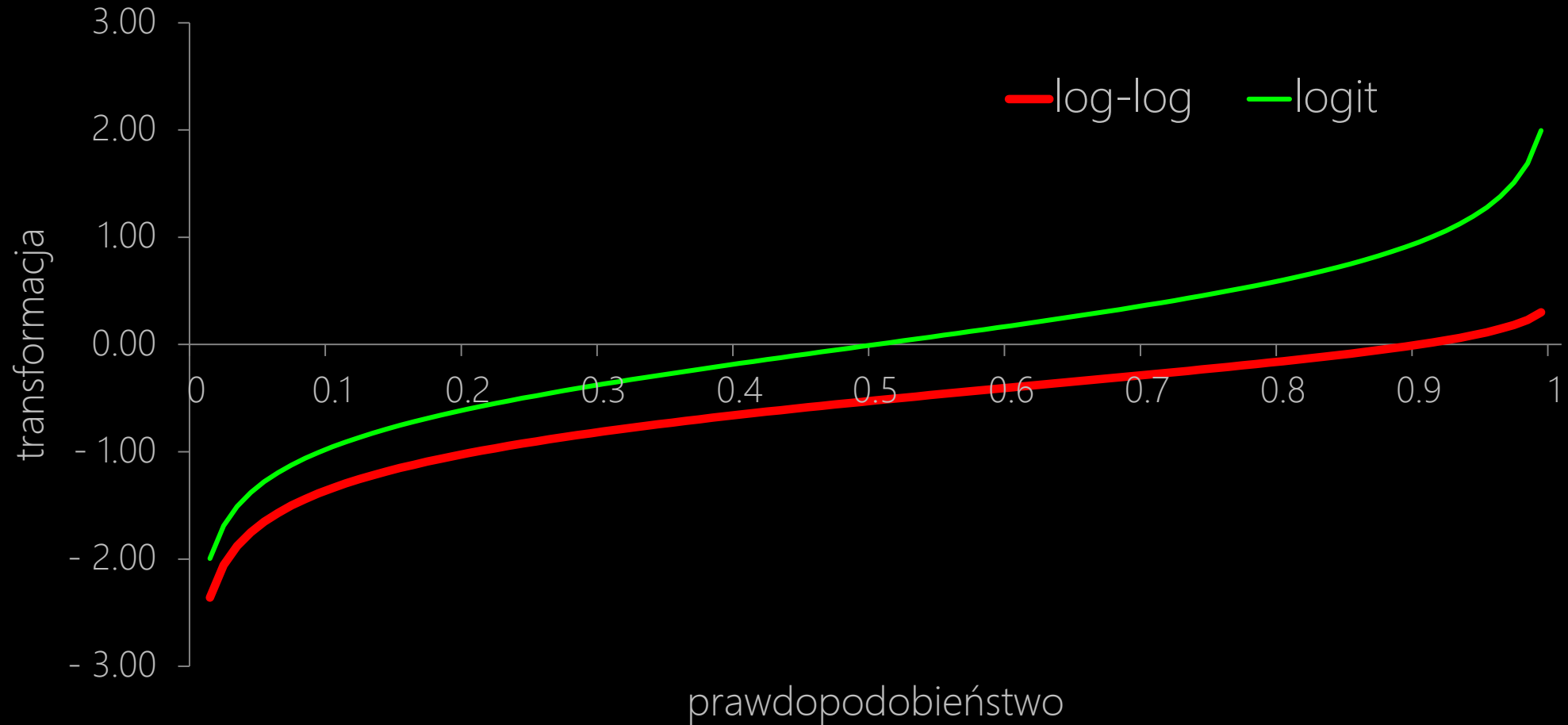
# Transformacja danych

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



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

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

$e \rightarrow \exp \rightarrow$  podstawa logarytmu naturalnego  $\approx 2,718$

# Estymacja parametrów

Iteracyjna metoda ważonych najmniejszych kwadratów → Fisher's method of scoring

pochodne dla  $\boldsymbol{\beta} = [\beta_0 \quad \beta_1]$

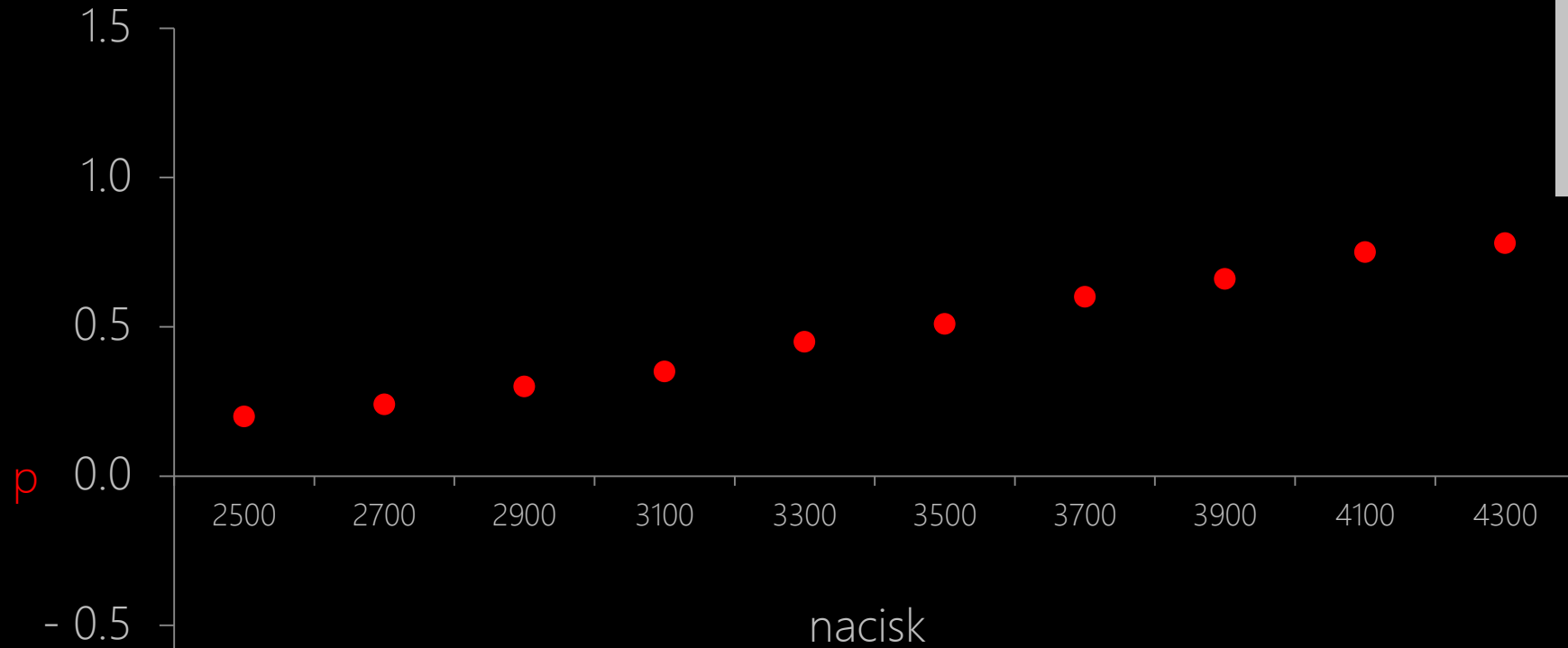
$$\boldsymbol{\beta}^{t+1} = \boldsymbol{\beta}^t + \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \left[ \binom{n}{k} p^{(t)k} (1 - p^{(t)})^{(n-k)} \right]}{-E \left( \frac{\partial^2 \left[ \binom{n}{k} p^{(t)k} (1 - p^{(t)})^{(n-k)} \right]}{\partial \beta_0 \beta_1} \right)}$$

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

Fisher's information matrix

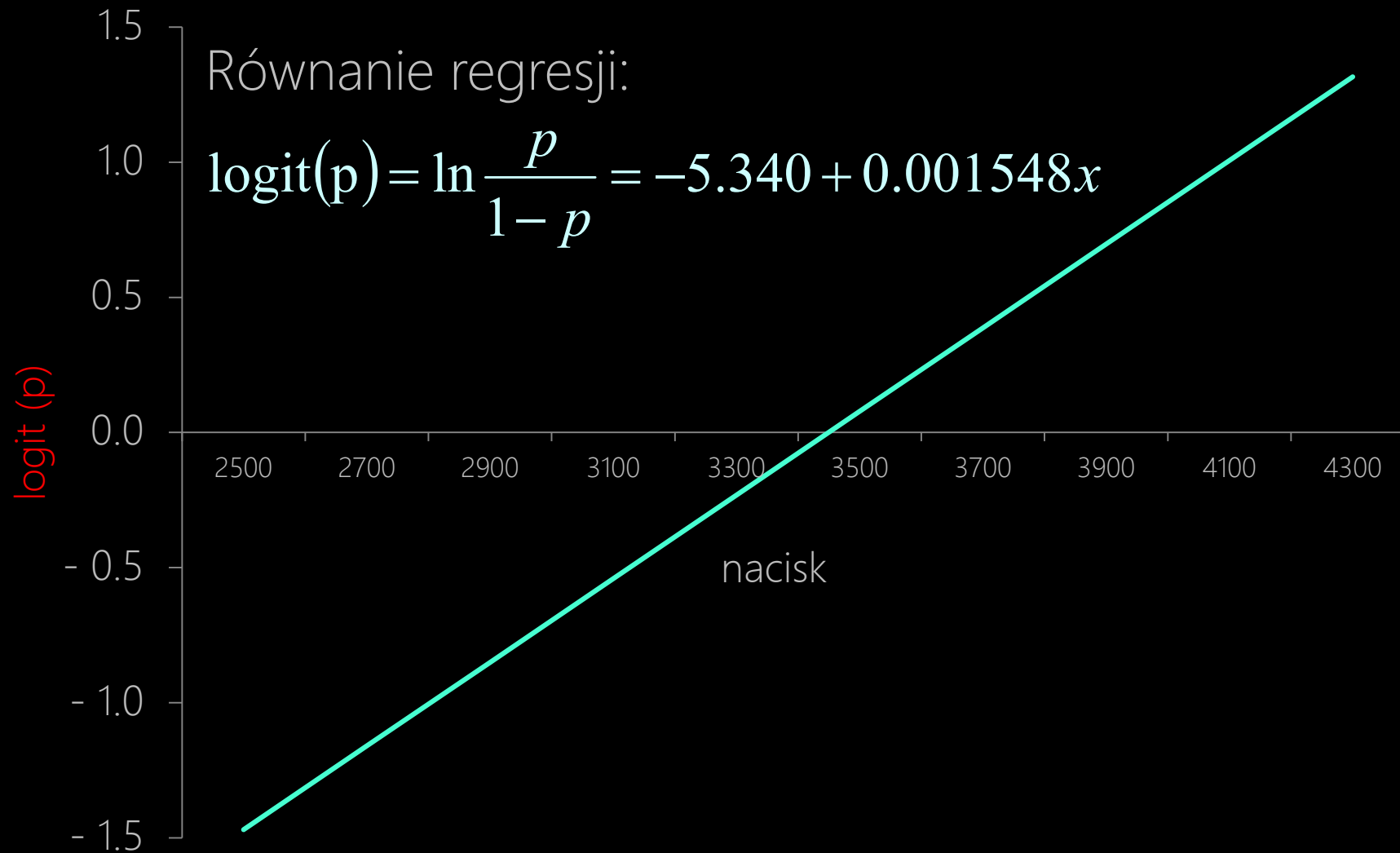
PRZYKŁAD

# Przykład - próba danych 1.

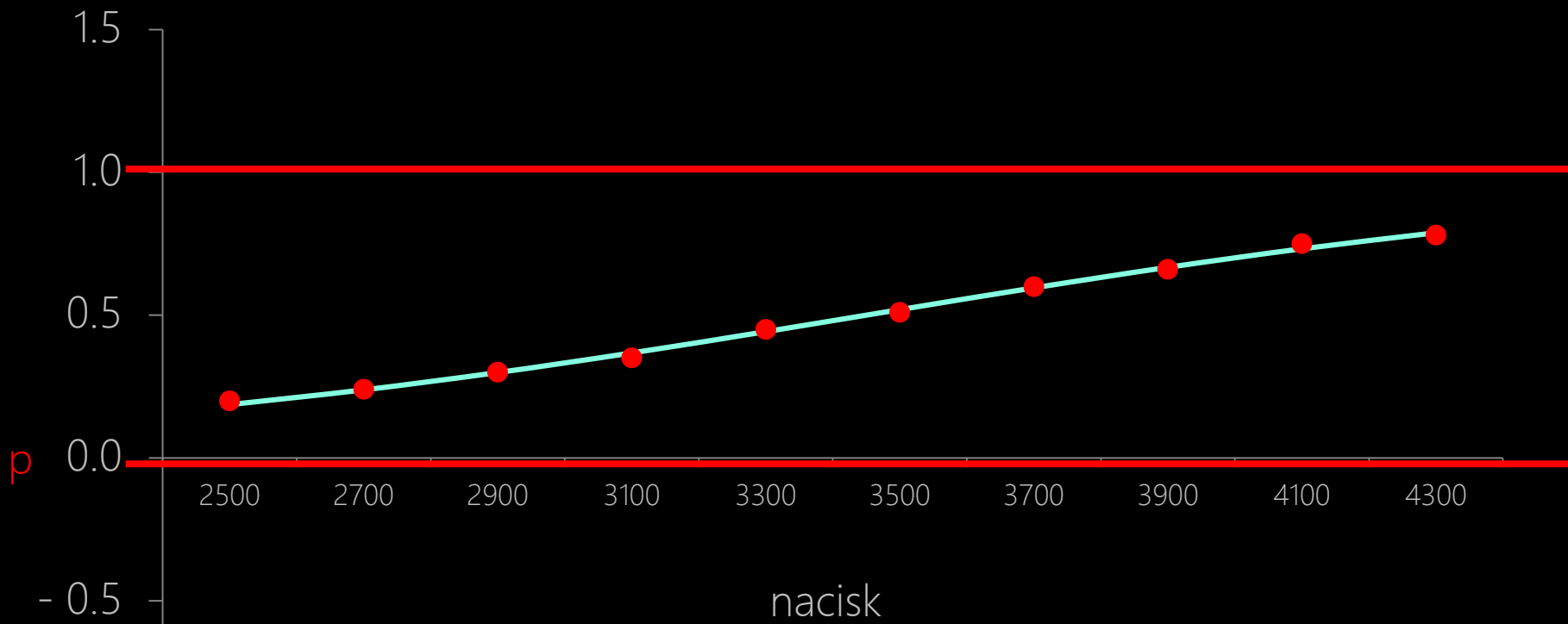


1. Prawdopodobieństwo uszkodzenia elementów metalowych w samolotach
2. Zmienna zależna  $[0,1]$
3. Nieliniowa zależność między  $x$  i  $y$

# Przykład - próba danych 1.



# 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 k=0/1 n=1	grupa x=0/1
1 	0=b.
1	0
0	0
...	...
1 	1=k.
1	1
0	1



1. Równanie regresji

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

2. Prawdopodobieństwo zachorowania jest mniejsze w grupie 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:

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

# Prediction of independence in bowel function after

## spinal cord injury: validation of a logistic regression

mc

**Objectives:** Recently, logistic regression models were developed to predict independence in bowel function 1 year after spinal cord injury (SCI) on a multicenter European SCI (EMSCI) dataset. Here, we evaluated the external

[Oma](#)

dataset.

$$P = \frac{e^f}{1 + e^f}, \text{ where } f = \beta_1 + \beta_2 * M_{tot} + \beta_3 * SCIM3a \quad 1$$

[Spinc](#)

**Setting:** Twenty-five

841

**Methods:** Two logistic regression models were developed using data from 277 patients derived from the EMSCI dataset. The external validation was evaluated using the EMSCI dataset utility. Discriminative ability was evaluated using the area under the curve (AUC) respectively, while c

where  $\beta_1 = -2.25046$ ,  $\beta_2 = 0.0486938$ , and  $\beta_3 = 0.4178468$  are constants. Note that SCIM3a represents independence in upper body dressing at baseline (SCIM subscore 3a) while  $M_{tot}$  represents the baseline total motor score, calculated by summing the motor scores for the upper extremity muscle groups and the lower extremity muscle groups. In addition to evaluating the full model, we analyzed the simplified model used by Pavese et al. [10]. In the simplified model, the SCIM3a term was removed; otherwise, it was the same as the full model. The simplified model is represented via the following equation:

$$P = \frac{e^f}{1 + e^f}, \text{ where } f = \beta_1 + \beta_2 * M_{tot} \quad 2$$



1. Rozkład dwumianowy - przykład
2. Transformacje danych dwumianowych
  - logit
  - probit
  - log-log
3. Równanie regresji logistycznej
  - równanie
  - estymacja parametrów równania
  - przykład