Historia Rachunku Prawdopodobieństwa i Statystyki
WYKŁAD VII: Ronald Fisher

MiNI PW, semestr zimowy

Ronald Fisher (1890-1962)

Studiował matematykę i astronomię w Cambridge.
Od 1919 r statystyk w Rothamsted Experimental Station, gdzie stworzył znany departament statystyki.
W 1933 w University College Karl Pearson przechodzi na emeryturę i kierowany przez niego Departament Statystyki zostaje podzielony na dwa:

- Departament Statystyki (kierownik: Egon Pearson);
- Departament Eugeniki (kierownik: Ronald Fisher);

1943-1957 profesor genetyki na Uniwersytecie w Cambridge.
1957-1960 Research Fellow Uniwersytetu w Adelaidzie.
Nie byl nigdy profesorem statystyki.

Głowne osiągnięcia

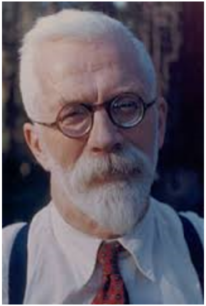
Książka: *Statistical Methods for Research Workers* (1925) i wiele następnych uzupełnianych wydań.

- podstawy statystyki teoretycznej (rodzina parametryczna, dostateczność, ..)
- planowanie eksperymentu;
- metoda randomizacji;
- metoda największej wiarogodności i jej optymalność;
- testy dla małych prób;
- studentyzacja i jej wykorzystanie;
- ......

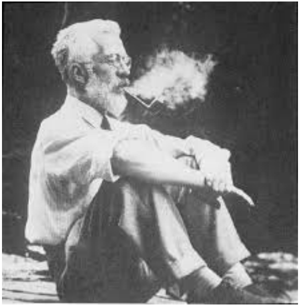
Ronald Fisher (1890-1962)

Ronald Fisher (1890-1962)

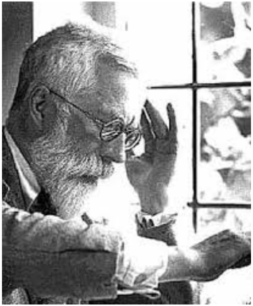
Ronald Fisher (1890-1962)



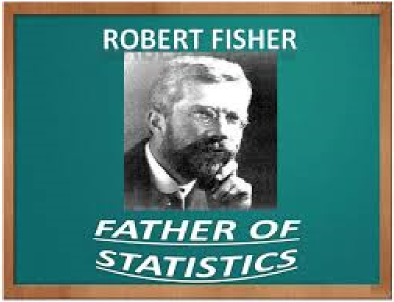
Ronald Fisher (1890-1962)



Ronald Fisher (1890-1962)



Ronald Fisher (1890-1962)



Rothamsted Experimental Station

Główne spory

- Fisher versus Pearson (miał rację);
- Fisher versus Mendel (miał rację);
- Fisher versus Neyman (?);
- Fisher versus Jeffreys (?);
- Fisher versus Barnard (nie miał).

Liczba stopni swobody i statystyka χ^2

$$\chi_1^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{n_i^2 + n^2 p_i^2 - 2nn_i p_i}{np_i} = \sum_{i=1}^{k} \frac{n_i^2}{np_i} - n.$$

$$\chi_2^2 = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = \sum_{i=1}^{k} \frac{n_i^2}{n\hat{p}_i} - n.$$

Podejście parametryczne Fishera:
$p_i = p_i(\theta)$, $\hat{p}_i = p_i(\hat{\theta})$.

Liczba stopni swobody i statystyka χ^2

Rozwiniecie $\chi_1^2 - \chi_2^2$ wykorzystujace ($\Delta\theta = \theta - \hat{\theta}$)

$$\frac{1}{p_i(\theta)} - \frac{1}{p_i(\hat{\theta})} = -\frac{p_i'(\hat{\theta})}{p_i^2(\hat{\theta})}\Delta\theta + \frac{1}{2}\Big(\frac{2}{(p_i^2(\hat{\theta})^3}[p_i'(\hat{\theta})]^2 - \frac{1}{p_i^2(\hat{\theta})}p_i''(\hat{\theta})\Big)(\Delta\theta)^2$$

Daje w efekcie

$$\chi_1^2 - \chi_2^2 \approx \frac{(\theta - \hat{\theta})^2}{\sigma_{\hat{\theta}}^2}.$$

Testy niezależności

W szczególności dla testów niezależności w tablicy kontyngencji $I \times J$ liczba stopni swobody wynosiła $(I-1) \times (J-1)$ a nie $I \times J - 1$ jak twierdził K. Pearson.
Fisher (1926): analiza 11688 tablic 2×2 wygenerowanych przez E. Pearsona dla niezależnych cech binarnych.
Srednia wartość statystyki chi kwadrat wyniosła $1.00001 \approx E(\chi_1^2)$ a nie $\approx 3 = E(\chi_3^2)$

Dane Mendla

Mendel analizował zależność koloru od genotypu:
groszek zielony: aa (a-gen recesywny) 25%
groszek żółty: aA i A (A-gen dominujący) 75%
R. Fisher połączył wyniki 84 niezależnych wyników eksperymentów Mendla (tablice 2×2). Sumaryczny $\chi^2 = 42$.

$$\chi^2_{84} \sim (84, 2 \times 84)$$

i $\sqrt{(}2 \times 84) \approx 13$.

$$P(\chi^2_{84} \leqslant 42) = 0.99996$$

Tylko w 4 przypadkach na 100 000 mozna otrzymać lepszy wynik przy spełnieniu hipotezy zerowej !

Metoda największej wiarogodności

Próbował udowodnić, że estymator NW jest optymalny (ma najmniejszą wariancję) wykorzystując to, że jest statystyką dostateczną.
Rozumowanie: S i T dwie statystyki dla parametru $\theta \in \Theta$. T- jest dostateczna. Załóżmy, że

$$(S, T) \sim N(\theta, \theta, \sigma_S, \sigma_T, \rho).$$

Wtedy

$$E(S|T = t) = \theta + \rho \frac{\sigma_S}{\sigma_T}(t - \theta)$$

Jeśli T- dostateczna, to prawa strona nie zależy od t. Zatem

$$\rho \frac{\sigma_S}{\sigma_T} = 1$$

i stąd

$$\sigma_T = \rho \sigma_S \leqslant \sigma_S.$$

T ma minimalną wariancję ! Problemy:
Estymator NW może nie istnieć, może nie być statystyką dostateczną ..