

ANOVA ZE ZMIENNĄ TOWARZYSZĄCĄ

Niech X zmienna, o której wiemy, że w grupie i -tej ($i=1,2,\dots,I$) ma rozkład $N(\mu_i, \sigma)$.

Analiza ANOVA bada, czy któraś ze średnich jest różna od pozostałych (hipoteza H_1).

Chcemy zbadać czy na te średnie ma wpływ zmienna towarzysząca Y , która w grupie i -tej ma średnią τ_i .

Jak wiadomo, miarą zależności między dwiema zmiennymi losowymi jest kowariancja.

Kowariancja próbkowa między X i Y nie może być zazwyczaj bezpośrednio obliczona, gdyż

Znamy jedynie oszacowania τ_i . Dlatego też hipoteza o wpływie zmiennej towarzyszącej Y na wartości X formułujemy jako hipotezę o wpływie średnich grupowych X i średnich grupowych Y ¹.

Czyli:

Testujemy hipotezę

H_0 : kowariancja w grupach X i $Y = 0$
przeciwko hipotezie

H_1 : kowariancja w grupach X i Y jest różna od 0

Statystyka testowa ma oczywiście postać

$$\frac{1}{I} \sum_{i=1}^I (\bar{X}_i - \bar{X})(\bar{Y}_i - \bar{Y})$$

Łatwo zobaczyć, że z dokładnością do stałego mnożnika $1/I$ jest ona równa

$$\sum_{i=1}^I \bar{X}_i (\bar{Y}_i - \bar{Y})$$

A więc możemy testować hipotezę

H_0 : $\sum_{i=1}^I \mu_i (\bar{Y}_i - \bar{Y}) = 0$
przeciwko

H_1 : $\sum_{i=1}^I \mu_i (\bar{Y}_i - \bar{Y}) \neq 0$

A więc, że kombinacja liniowa $\sum_{i=1}^I \mu_i C_i = 0$ z $C_i = \bar{Y}_i - \bar{Y}$ ($\sum_{i=1}^I C_i = 0$)

W przykładzie o rybach

i	1	2	3	4	5	6	Średnia
$\bar{X}_i$	56,41	60,89	62,43	67,00	64,21	63,34	62,13
$C_i = \bar{Y}_i - \bar{Y}$	2,5	-1,5	0,5	1,5	-4,5	1,5	32,5
n_i	16	14	17	14	9	14	

Efekt: $g = \sum_{i=1}^I \bar{X}_i (\bar{Y}_i - \bar{Y}) = -12,53$

Odchylenie standardowe łączne X -ów wynosi 15,46, więc

$$SE(g) = 15,46 \sqrt{\frac{2,5^2}{16} + \frac{1,5^2}{14} + \frac{0,5^2}{17} + \frac{1,5^2}{14} + \frac{4,5^2}{9} + \frac{1,5^2}{14}} = 27,38$$

Statystyka t-Studenta z 78 ss $-12,53/27,38 = -0,458$ p-value jednostronne $p=0.32$ więc nie ma związku z rozmiarami ciała.

¹ Jest to równoważne hipotezie, że *wewnątrz* grup zależności między X i Y nie występują, co jest logiczne.