ANALIZA KOWARIANCJI -- ZMIENNA TOWARZYSZĄCA

Analiza kowariancji to często stosowana metoda statystyczna łącząca w sobie elementy analizy wariancji, korelacji i regresji. Główny cel metody jest podobny do analizy wariancji: dać odpowiedź czy analizowany czynnik doświadczalny wpływa w sposób istotny na badaną cechę. Tak więc metodę tę stosujemy gdy na wybraną cechę działa czynnik doświadczalny na kilku poziomach. Dodatkowo, zaletą metody analizy kowariancji jest możliwość wyeliminowania wpływu innej cechy mającej wpływ na cechę badaną. Tą inną cechę nazywać będziemy zmienną towarzyszącą.

PRZYKŁAD ZASTOSOWANIA

Wydajność mleczna krowy zależy między innymi od takich czynników jak jej wiek wycielenia i sezon wycielenia. Stosując analizę wariancji w taki sposób, że czynnikiem doświadczalnym jest sezon wycielenia, a wydajność mleczna badaną cechą, nie uwzględnia się faktu wpływu wieku na tę cechę. Fakt ten będzie oczywiście miał wpływ na wyniki porównania. Zatem należy „pozbyć się" wpływu wieku na wyniki obliczeń. Zakładamy więc, iż chcemy zbadać wpływ sezonu wycielenia na mleczność, natomiast wiek krowy traktujemy jako zmienną towarzyszącą i jej wpływ musimy wyeliminować. W tym celu musimy zmierzyć wartości dwóch cech:
- mleczność - będziemy ją porównywali,
- wiek krowy - jej wpływ wyeliminujemy - zmienna towarzysząca.

MODEL LINIOWY

W metodzie analizy kowariancji najprostszy model przedstawia się następująco:

$$y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \overline{x}) + e_{ij} \ ,$$

gdzie:
α - czynnik doświadczalny,
β - współczynnik regresji między zmiennymi,
y - cecha badana
x - zmienna towarzysząca
pozostałe elementy modelu jak w analizie wariancji.

ZAŁOŻENIA

Stosowanie metody analizy kowariancji ma sens tylko wtedy gdy istnieje zależność między badaną cechą a zmienną towarzyszącą. Znaczy to, że gdy nie ma takiej zależności wystarczy zastosować prostszą metodę (jaką ?).

Niezbędne jest aby spełnione zostały następujące założenia w stosunku do zmiennej niezależnej x i zależnej y:

- zmienna y jest zmienną losową zależną od x o rozkładzie normalnym i jednakowych wariancjach dla każdego x
- zmienna towarzysząca x ma stałą wariancję i nie obserwuje się dla niej istotnych różnic między średnimi obiektowymi
- istnieje zależność liniowa między y i x w każdej grupie (obiekcie)
- współczynnik regresji jest jednakowy dla każdego obiektu

TABELA ANALIZY KOWARIANCJI

zmienność	st.swob.	sumy kwadratów i iloczynów			odchylenie od regresji		
		SS_{xx}	SS_{yy}	SSxy	st.swob.	SS	MS
Międz.ob.	t-1	T_{xx}	T_{yy}	T_{xy}	t-2	T	
Wew.ob.	n-t	E_{xx}	E_{yy}	E_{xy}	n-t-1	E	MS_E
całkowita	n-1	G_{xx}	G_{yy}	G_{xy}	n-2	G	
suma kwadratów do testowania poprawki średnich					t-1	$T'=G-E$	$MS_{T'}$
Reszta					1	D	

PRZYKŁAD

Porównywano masę ciała bażantów łownych z trzech linii genetycznych: polskiej, francuskiej 312 i 112 uwzględniając przy tym ich wiek. Zanotowano następujące obserwacje u pięciu losowo wybranych sztuk z każdej linii. Sprawdź czy linie różnią się masą ciała zwierząt.

	linia polska	linia francuska 312	linia francuska 112	
1	562 10	550 10	420 9	
2	810 13	645 11	561 11	
3	891 14	868 13	620 12	
4	1072 17	1023 15	778 14	
5	1227 19	1217 18	820 15	
sumy	4562 73	4303 67	3199 61	12064 201
średnie	912.4 14.6	860.6 13.4	639.8 12.2	804.3 13.4

zmienność	st.swob.	sumy kwadratów i iloczynów			odchylenie od regresji		
		SS_{xx}	SS_{yy}	SSxy	st.swob.	SS	MS
między obiektami	2	14.4	209577.7	1635.6	1	23800.8	
wewnątrz obiektów	12	113.2	661059.2	8580.8	11	10616.4	965.1
całkowita	14	127.6	870636.9	10216.4	13	52652.3	
suma kwadratów do testowania średnich					2	42035.9	21018.0
reszta					1	18295.1	