ANALIZA WARIANCJI

Analiza wariancji to powszechnie stosowana metoda statystyczna pozwalająca na ocenę istotności różnic wielu średnich.

ANALIZA WARIANCJI DLA KLASYFIKACJI POJEDYNCZEJ

Wcześniej poznany test pozwalał na ocenę istotności różnicy pary średnich. Jeśli mamy do czynienia z więcej niż 2 średnimi musimy zastosować metodę analizy wariancji. Stawiamy wtedy następującą hipotezę:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_t$$

wobec hipotezy alternatywnej:

$$H_1: \text{średnie różnią się}$$

Z sytuacją taką mamy do czynienia jeśli **czynnik doświadczalny** działa na więcej niż 2 **poziomach**: np. porównujemy wpływ 3 różnych pasz, 4 rasy zwierząt, t sposobów utrzymania itp.

Rozważmy sytuację kiedy czynnik doświadczalnym działą na t poziomach. Sprawdzimy jak czynnik wpływa na wybraną **cechę - zmienną.** Zwierzęta biorące udział w doświadczeniu (**materiał doświadczalny**) wybiera się losowo i przydziela do t grup, tzw. **obiektów doświadczalnych**. W takiej sytuacji każde zwierzę stanowi **jednostkę doświadczalną** bowiem dostarcza pojedynczej informacji (w innych sytuacjach jednostką może być grupa zwierząt np. jeśli rozpatrujemy masę miotu jednostkę doświadczalną stanowią wszystkie prosięta od jednej lochy). Zwierzęta wchodzące w skład jednego obiektu doświadczalnego poddane są działaniu czynnika doświadczalnego na jednakowym poziomie.

Obiekt 1	Obiekt 2	Obiekt t
y_{11}	y_{21}	y_{t1}
y_{12}	y_{22}	y_{t2}
y_{13}	y_{2n2}	y_{t3}
y_{1n1}		y_{tnt}

Pierwszy wskaźnik przy y oznacza przynależność do i-tego obiektu doświadczalnego, natomiast drugi oznacza kolejne zwierze - **replikację** w obiekcie. W t-tym obiekcie mamy nt replikacji.

ZAŁOŻENIA METODY ANALIZY WARIANCJI

Stosowanie metody analizy wariancji (anova - analysis of variance) wymaga aby analizowana cecha miała rozkład normalny oraz aby wariancje obiektowe były jednorodne tzn. nie różniły sie istotnie:

$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2$$

TABELA ANALIZY WARIANCJI dla klasyfikacji pojedynczej

Zanim zweryfikuje się hipotezę zerową należy wypełnić tabelę analizy wariancji.

źródło zmienności	stopnie swobody	sumy kwadratów (SS)	średnie kwadraty (MS)	wartość statystyki F
między obiektami	$\nu_1 = t-1$	SS_t	MS_t	$F= MS_t / MS_e$
wewnątrz obiektów (błąd)	$\nu_2 = n.-t$	SS_e	MS_e	

gdzie: n**.** - suma wszystkich oserwacji;

$$SS_t = \sum_{i=1}^{t}\left(\bar{y}_i - \bar{\bar{y}}\right)^2 n_i, \qquad SS_e = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2,$$

$\bar{\mathbf{y}}_\mathbf{i}$ - wartość średnia z i-tego obiektu, $\bar{\bar{\mathbf{y}}}$ - średnia ze wszystkich obserwacji.

Do weryfikacji hipotezy zerowej używa się testu F (Fischera-Snedecora). Obliczoną wartość F porównuje się z F dla określonego α i ν_1, ν_2 stopni swobody.

MODEL LINIOWY

Każdą obserwację możemy opisać przy użyciu tzw. modelu liniowego:

$$\mathbf{y_{ij} = \mu + \tau_i + e_{ij}}$$

gdzie y_{ij} oznacza j-tą obserwację z i-tego obiektu, μ – średnią wartość cechy w populacji, τ_i efekt i-tego obiektu, natomiast e_{ij} błąd czyli efekt związany ze zmiennością osobniczą jak i błąd pomiaru. Jest to najprostszy model liniowy bowiem występuje w nim tylko jeden czynnik doświadczalny (τ_i). Modele liniowe mogą być rozbudowane o dalsze czynniki doświadczalne jak i o interakcję tych czynników. W zależności od tego czy rozpatrywane efekty są losowe czy stałe model nazywa się modelem losowym, stałym lub mieszanym.

✍ PRZYKŁAD

Porównywano wydajność mleczną 3 ras krów w szczytowym okresie laktacji. U losowo wybranych sztuk zanotowano następujące obserwacje:

ncb: 35, 40, 32, 40, 41, 39
hf: 35, 48, 42, 48, 45
nczb: 32, 39, 42, 32, 35, 40

Stąd: $n_1 = 6$, $n_2 = 5$, $n_3 = 6$

źródło zmienności	stopnie swobody	sumy kwadratów (SS)	średnie kwadraty (MS)	wartość statystyki F
między obiektami	2	146,4	73,2	F= 3,776
wewnątrz obiektów (błąd)	14	271,4	19,4	

n. = 17, $\overline{y}_1 = 37{,}8$ $\overline{y}_2 = 43{,}6$ $\overline{y}_3 = 36{,}7$.

W tabelach znajdujemy wartości F dla poziomu istotności 0.05 i odpowiednich stopni swobody → 3.739. W związku z tym, że $F_{obl} > F_{0.05}$ odrzucamy hipotezę zerową ➔ stwierdzamy, że średnie wydajności krów w szczytowym okresie laktacji różnią się istotnie ➔ a to oznacza, że rasa wpływa istotnie na wydajność dzienną krów w szczytowym okresie laktacji.

ZADANIA

① Porównywano użytkowość wełnistą owiec z trzech rejonów kraju: Podhala, Wielkopolski i Mazur. U losowo wybranych sztuk zważono runo:

Podhale	4,5	4	3	3,5	5	6,5
Wielkopolska	4.5	5	7	5	3.5	7,5
Mazury	5	6	5.5	3,5	3	

Sprawdź czy owce z różnych rejonów kraju różnią się użytkowością wełnistą.

② W zakładzie doświadczalnym testowano 4 linie towarowe brojlerów. Po 6 tygodniach tuczu wylosowano po kilka sztuk z 4 linii i porównano masy ciała. Przeprowadź stosowne obliczenia i daj odpowiedź.

1 linia -	1800	1650	1900	1700	1750	1650	1700
2 linia -	1800	1800	1850	2000	2050	1950	1950
3 linia -	1800	1900	1750	1800	2000	1950	1950
4 linia -	1800	1600	1950	1700	1800	1650	1700