Analiza wariancji (ANOVA)

Jednoczynnikowa ANOVA

- Na test jednoczynnikowej analizy wariancji możemy patrzeć jak na uogólnienie testu t-Studenta dla dwóch prób niezależnych, na przypadek $k, (k > 2)$ prób niezależnych
- Model:

  $$X_{ij} = \mu_i + \varepsilon_{ij}, \qquad i = 1, \dots, k, \qquad j = 1, \dots, n_i$$

  gdzie:
  μ_i - "prawdziwa" wartość badanej cechy w i-tej grupie
  ε_{ij} - błędy (niezależne zmienne losowe o jednakowym rozkładzie $N(0, \sigma^2)$)

Jednoczynnikowa ANOVA

- **Hipoteza zerowa:**
  Wartości oczekiwane (średnie) badanej cechy w k grupach **nie różnią się istotnie**:
  $$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_k$$
- **Hipoteza alternatywna:**
  Co najmniej dla jednej pary grup, wartości oczekiwane (średnie) badanej cechy **różnią się istotnie**:
  $$H_1: \quad \sim H_0$$
- **Statystyka testowa:**
  $$F = \frac{n-k}{k-1}\,\frac{SSA}{SSE}$$

Jednoczynnikowa ANOVA

- gdzie

$$SSA = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2 \qquad SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2$$

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} \qquad n = \sum_{i=1}^{k} n_i$$

- **Rozkład statystyki testowej:**

$$F\Big|_{H_0} \sim F(k-1, n-k)$$

Tabela analizy wariancji

- Tradycyjnie wyniki analizy wariancji przedstawiamy w postaci tabeli:

Zmienność	Suma kwadratów	Liczba stopni swobody	Średnie kwadraty	Statystyka testowa
Pomiędzy grupami	SSA	$k-1$	$MSA = \frac{SSA}{k-1}$	F
Wewnątrz grup	SSE	$n-k$	$MSE = \frac{SSE}{n-k}$	
Całość	SST	$n-1$	$MST = \frac{SST}{n-1}$	

Założenia jednoczynnikowej analizy wariancji

- Niezależność obserwacji dla poszczególnych jednostek eksperymentalnych
- Błędy mają rozkłady normalne z zerową wartością oczekiwaną (brak błędu systematycznego) i jednorodną wariancję

- **Uwaga:**
  Założenie jednorodności wariancji możemy zweryfikować testem Barletta

Test Barletta

- **Założenia:** Model normalny, wiele prób niezależnych
- **Hipoteza zerowa:**

$$H_0: \quad \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

- **Hipoteza alternatywna:**

$$H_1: \ \sim H_0$$

- **Statystyka testowa:**

$$B = \frac{1}{C}(n-k)\ln MSE - \sum_{i=1}^{k}(n_i - 1)\ln S_i^2$$

$$C = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{n-k}\right)$$

- **Rozkład statystyki testowej:**

$$B\,\Big|_{H_0} \sim \chi^2(k-1) \quad \text{(graniczny)}$$

Porównania wielokrotne (post hoc)

- Procedury porównań wielokrotnych stosujemy wtedy, gdy zostanie odrzucona hipoteza zerowa w analizie wariancji
- **Procedura NIR – Fishera**
- Polega na testowaniu, dla każdej pary $(i, j), i, j = 1,2, \dots, k, i \neq j$ oddzielnie hipotezy zerowej:

$$H_0: \quad \mu_i = \mu_j$$

  przeciwko hipotezie alternatywnej:

$$H_1: \quad \mu_i \neq \mu_j$$

- Wartość statystyki testowej:

$$t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MSE}} \sqrt{\frac{n_i n_j}{n_i + n_j}}$$

- Rozkład statystyki testowej (t-Studenta):

$$t\Big|_{H_0} \sim t(n-k)$$

Porównania wielokrotne (post hoc)

- **Procedura HSD – Tukey'a**
- **Założenie:**

$$n_1 = n_2 = \cdots = n_k = m$$

- Polega na testowaniu, jednocześnie dla wszystkich par $(i, j), i, j = 1,2, \ldots, k, i \neq j$ hipotez zerowych:

$$H_0: \quad \mu_i = \mu_j$$

  przeciwko hipotezom alternatywnym

$$H_1: \quad \mu_i \neq \mu_j$$

- **Wartość statystyki testowej:**

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MSE}} \sqrt{m}$$

- **Rozkład statystyki testowej** ma rozkład q (rozkład studentyzowanego rozstępu) z k i $m - k$ stopniami swobody

$$q \Big|_{\mathrm{H_0}} \sim q(k, m-k)$$