

Analiza wariancji

Weronika Przysiężna

Marzec 2025

1 Teoretyczne wprowadzenie do analizy wariancji

1.1 Wstęp

Analiza wariancji jest analizą statystyczną, która wykrywa różnice między dwiema lub więcej grupami określonymi dla pojedynczego czynnika lub zmiennej niezależnej. Identyfikuje ona zmienność lub wariancję pomiędzy obserwacjami przypisując ją różnym źródłom, które (po odpowiednim przetestowaniu) wskazują, czy zaobserwowane różnice między średnimi grupowymi są prawdopodobnie rzeczywiste, czy jedynie wynikiem przypadku. [2]

W naszym badaniu skupimy się na wykrywaniu różnic przy pomiarach na różnych osobach.

1.2 Założenia dotyczące jednoczynnikowej analizy wariancji (ANOVA) z jednym czynnikiem międzyobiektywowym [1]

- **Zmienna zależna mierzona na skali ilościowej:** zmienna zależna powinna być zmienną ilościową (na poziom interwałowym lub ilorazowym).
- **Losowość i niezależność obserwacji:** nie ma związku między obserwacjami w każdej grupie lub między samymi grupami, a w każdej grupie są różni uczestnicy badania i żaden uczestnik nie należy do więcej niż jednej grupy; uczestnicy badania są dobierani losowo.
- **Równoliczność obserwacji w grupach:** poszczególne kategorie zmiennej niezależnej powinny być statystycznie równoliczne (aby sprawdzić, czy analizowane grupy różnią się istotnie statycznie pod względem liczebności, można zastosować test zgodności Chi-kwadrat).
- **Rozkład normalny:** rozkład wyników w analizowanych grupach jest zbliżony do rozkładu normalnego (oceny tego założenia można dokonać stosując test Kołomogorowa-Smirnova lub Shapiro-Wilka).
- **Wariancje w grupach są jednorodne (homogeniczność wariancji):** zmienność w każdej porównywanej grupie powinna być podobna; jeśli wariancje różnią się między grupami, to można zastosować test Welcha lub Browna-Forsythe'a, które wprowadzają poprawkę na nierówne wariancje do statystyki F.

1.3 Jednoczynnikowa ANOVA

1.3.1 Suma kwadratów SS (ang. Sum of Squares)

Wariancja próby mierzy zmienność w dowolnym zbiorze obserwacji poprzez obliczenie sumy kwadratów odchyłeń od ich średniej:

$$SS = \sum (X - \bar{X})^2.$$

Następnie suma kwadratów SS jest dzielona przez liczbę stopni swobody $n - 1$:

$$s^2 = \frac{SS}{df},$$

gdzie:

- \bar{X} - średnia próby,
- s^2 - wariancja próby,
- $df = n - 1$ - stopnie swobody.

1.3.2 Średnia kwadratów MS (ang. Mean Square)

Średnia kwadratów to oszacowanie wariancji uzyskane przez podzielenie sumy kwadratów SS przez liczbę stopni swobody $n - 1$. Ogólny wzór na oszacowanie wariancji ma postać:

$$MS = \frac{SS}{df},$$

gdzie:

- MS - średnia kwadratów,
- SS - suma kwadratów odchyłeń od średniej,
- $df = n - 1$ - liczba stopni swobody.

1.3.3 Wzory definicyjne na sumy kwadratów [2]:

1. SS_{total} - całkowita suma kwadratów odchyłeń od średniej ogólnej (zmienność całkowita)

$$SS_{total} = \sum (X - \bar{X}_{grand})^2.$$

Równoważny wzór obliczeniowy:

$$SS_{total} = \sum X^2 - \frac{G^2}{N}.$$

2. $SS_{between}$ - suma kwadratów odchyłeń średnich grupowych od średniej ogólnej (zmienność między grupami)

$$SS_{between} = \sum n(\bar{X}_{group} - \bar{X}_{grand})^2.$$

Równoważny wzór obliczeniowy:

$$SS_{between} = \sum \frac{T^2}{n} - \frac{G^2}{N}.$$

3. SS_{within} - suma kwadratów odchyłeń indywidualnych wyników w grupie od średnich grupowych (zmienność wewnątrz grup)

$$SS_{within} = \sum (X - \bar{X}_{group})^2.$$

Równoważny wzór obliczeniowy:

$$SS_{within} = \sum X^2 - \sum \frac{T^2}{n}.$$

4. Sprawdzamy dokładność obliczeniową weryfikując równość:

$$SS_{total} = SS_{between} + SS_{within}$$

Oznaczenia:

- X - pojedyncza wartość obserwowana,
- \bar{X}_{group} - średnia dla danej grupy,
- \bar{X}_{grand} - średnia ogólna dla całej próby,
- T - suma wartości w grupie,
- n - liczba obserwacji w grupie,
- G - suma wartości dla wszystkich grup (suma ogólna),
- N - całkowita liczba obserwacji (sumaryczna wielkość próby).

1.3.4 Stopnie swobody (df):

1. $df_{total} = N - 1$,
2. $df_{between} = k - 1$,
3. $df_{within} = N - k$,

gdzie:

- N - całkowita liczba obserwacji (sumaryczna wielkość próby),
- k - liczba grup.

Sprawdzamy dokładność obliczeń weryfikując równość:

$$df_{total} = df_{between} + df_{within}.$$

1.3.5 Wzory na średnie kwadratów [2]:

1. $MS_{between}$ - średni kwadrat odchyłeń między grupami (zmienność między średnimi dla grup)

$$MS_{between} = \frac{SS_{between}}{df_{between}}$$

2. MS_{within} - średni kwadrat odchyłeń wewnątrz grupy (zmienność wyników wewnątrz grupy; mierzy jedynie błąd losowy)

$$MS_{within} = \frac{SS_{within}}{df_{within}} = MS_{error}$$

1.3.6 Rozkład F-Snedecora

Liczmy statystykę testową F jako:

$$F = \frac{MS_{between}}{MS_{within}}$$

Określamy obszar krytyczny jako:

$$Q = \{F : F \geq F_\alpha\},$$

gdzie F_α jest wartością krytyczną odczytaną z tablic rozkładu F-Snedecora dla $(df_{between}, df_{within})$ stopni swobody, czyli $F(df_{between}, df_{within})$.

1. Jeżeli $F \in Q$ ($F \geq F_\alpha$), to odrzucamy hipotezę zerową H_0 na korzyść hipotezy alternatywnej H_A i wnioskujemy, że badane średnie nie są sobie równe.
2. Jeżeli $F \notin Q$ ($F < F_\alpha$), to nie ma podstaw do odrzucenia hipotezy zerowej H_0 i wnioskujemy, że badane średnie są równe.

1.3.7 Testy post-hoc (test HSD Tukey'a)

1.3.8 Obliczenie siły efektu (d Cohen'a)

2 Bibliografia

References

- [1] Norm O'Rourke, Larry Hatcher, and Edward J. Stepanski. *A Step-by-Step Approach to Using SAS for Univariate and Multivariate Statistics*. SAS Institute, Cary, NC, USA, 2th edition, 2005.
- [2] Robert S. Witte and John S. Witte. *Statistics*. John Wiley & Sons, Hoboken, NJ, USA, 9th edition, 2010.