

Izračun variance razlike povprečji vzorca in podvzorca

Luka Štrlekar

Imamo vzorec n enot, kjer del vzorca pripada neki skupini (podvzorec n_p enot). Želimo preveriti statistično značilnost razlike povprečji med celotnim vzorcem (\bar{x}_s) in podvzorcem (\bar{x}_p).

Zapišimo najprej povprečje celotnega vzorca \bar{x}_s v obliki linearne kombinacije povprečji obeh podvzorcev ($w \in (0, 1)$ je delež vsebovanosti prvega podvzorca v celotnem vzorcu):

$$\bar{x}_s = w\bar{x}_{p1} + (1 - w)\bar{x}_{p2}$$

Razliko med povprečjem celotnega vzorca in podvzorca lahko izrazimo kot:

$$\begin{aligned}\bar{x}_{p1} - \bar{x}_s &= \bar{x}_{p1} - [w\bar{x}_{p1} + (1 - w)\bar{x}_{p2}] \\ &= \bar{x}_{p1} - w\bar{x}_{p1} - \bar{x}_{p2} + w\bar{x}_{p2} \\ &= (1 - w)(\bar{x}_{p1} - \bar{x}_{p2})\end{aligned}\tag{1}$$

, oziroma kot:

$$\begin{aligned}\bar{x}_{p2} - \bar{x}_s &= \bar{x}_{p2} - [w\bar{x}_{p1} + (1 - w)\bar{x}_{p2}] \\ &= \bar{x}_{p2} - w\bar{x}_{p1} - \bar{x}_{p2} + w\bar{x}_{p2} \\ &= w(\bar{x}_{p2} - \bar{x}_{p1})\end{aligned}\tag{2}$$

Razliko vzorca in podvzorca, ki nista neodvisna zaradi prekrivanja, smo prevedli na razliko dveh neodvisnih podvzorcev, pomnoženih s konstanto prekrivanja (w). Izračun variance razlike je zaradi neodvisnosti sedaj:

$$\begin{aligned}Var(\bar{x}_{p1} - \bar{x}_s) &= Var[(1 - w)(\bar{x}_{p1} - \bar{x}_{p2})] \\ &= (1 - w)^2 [Var(\bar{x}_{p1}) + Var(\bar{x}_{p2})]\end{aligned}\tag{3}$$

, oziroma:

$$\begin{aligned}Var(\bar{x}_{p2} - \bar{x}_s) &= Var[w(\bar{x}_{p2} - \bar{x}_{p1})] \\ &= w^2 [Var(\bar{x}_{p1}) + Var(\bar{x}_{p2})]\end{aligned}\tag{4}$$

, kjer sta $Var(\bar{x}_{p1})$ in $Var(\bar{x}_{p2})$ vzorčni varianci podvzorcev:

$$Var(\bar{x}_{p1}) = \frac{s_{p1}^2}{n_{p1}} \quad Var(\bar{x}_{p2}) = \frac{s_{p2}^2}{n_{p2}}$$

Simulacije

Preverimo teoretične izračune še s simulacijami. Pri vsaki iteraciji simulacije generiramo 1000 enot iz normalne porazdelitve ($N \sim (3, 1^2)$) in jih razdelimo v dva podvzorca skladno z w . Za vsak w izvedemo 10^4 simulacij.

Primer $\bar{x}_{p1} - \bar{x}_s$:

| w | Izračunana vrednost | Ocenjena vrednost (simulacije) |
|-----|---------------------|--------------------------------|
| 0.1 | 0.00900 | 0.00898 |
| 0.2 | 0.00400 | 0.00400 |
| 0.3 | 0.00233 | 0.00233 |
| 0.4 | 0.00150 | 0.00148 |
| 0.5 | 0.00100 | 0.00099 |
| 0.6 | 0.00067 | 0.00069 |
| 0.7 | 0.00043 | 0.00042 |
| 0.8 | 0.00025 | 0.00025 |
| 0.9 | 0.00011 | 0.00011 |

Primer $\bar{x}_{p2} - \bar{x}_s$:

| w | Izračunana vrednost | Ocenjena vrednost (simulacije) |
|-----|---------------------|--------------------------------|
| 0.1 | 0.00900 | 0.00898 |
| 0.2 | 0.00400 | 0.00400 |
| 0.3 | 0.00233 | 0.00233 |
| 0.4 | 0.00150 | 0.00148 |
| 0.5 | 0.00100 | 0.00099 |
| 0.6 | 0.00067 | 0.00069 |
| 0.7 | 0.00043 | 0.00042 |
| 0.8 | 0.00025 | 0.00025 |
| 0.9 | 0.00011 | 0.00011 |