Assignment 2 Q3

Fenglin Chen

27/10/2021

Table of Contents

Summary

- $a(\mathcal{P})$ is an attribute calculated over a population
- Influence is how much $a(\mathcal{P})$ changes when one unit u is removed
  - $u \in \mathcal{P}$
- Used to find interesting units

Math Formula

For $u \in \mathcal{P}$, the change in attribute value can be written as:

$$\Delta(a, u) = a(y_1, y_2, ..., y_{u-1}, y_u, y_{u+1}, ..., y_n) - a(y_1, y_2, ..., y_{u-1}, y_{u+1}, ..., y_n)$$

Math Example

The influence on the range of the population $(1, 3, 5, 7, 9)$ when 9 is removed is:

$$\begin{aligned} \text{influence} &= \text{range}(1, 3, 5, 7, 9) - \text{range}(1, 3, 5, 7) \\ &= 8 - 6 \\ &= 2 \end{aligned}$$

General Code

```r
influence <- function(attr) {
  delta = rep(0, length(pop))
  pop.attr <- attr(pop)
  for (i in 1:length(pop)) {
    delta[i] = pop.attr - attr(pop[-i])
  }
  return(delta)
}
```

Code Example

Influence on standard deviation:

```
pop <- 1:100
pop.infl <- influence(sd)

plot(pop.infl, pch=19, xlab="Unit", ylab="Influence",
     main="Effect on Standard Deviation")
```

Effect on Standard Deviation