Assignment 2 Q3

Fenglin Chen

27/10/2021

## Summary

When an attribute is calculated on a population, the influence of each unit is how much the attribute value changes when it’s removed from the population. This is different from sensitivity, where arbitrary values inside or outside the population are added to the group.

Influence is often used to locate interesting units in the population. For example, a unit with large influence (greatly alters the attribute value when removed) could match a particular filter well, have a very large/small value, or be an outlier. Whether the effect is good or bad, it’s important to be aware of influential units when making summary reports to be more representative.

## Mathematical Formula

For every unit , the change in attribute value can be written as:

Where unit is removed from the second attribute.

## Mathematical Example

For example, the influence on the range of the population when is removed is:

Since lowers the range by when it’s removed, it is an influential unit. On the other hand, values have no effect on the range.

## General Code

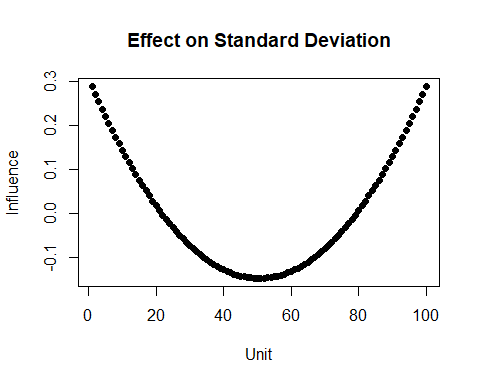
Below is a general formula in R. The function takes input attr: the attribute applied to the population.

influence <- function(attr) {  
 delta = rep(0, length(pop)) # Store the results, one for each unit  
 pop.attr <- attr(pop) # The attribute over the original population  
 for (i in 1:length(pop)) {  
 delta[i] = pop.attr - attr(pop[-i]) # Calculate on all units except one  
 }  
 return(delta)  
}

## Code Example

Test the general formula on the standard deviation of a population and plot the change caused by removing each unit:

pop <- 1:100  
pop.infl <- influence(sd)  
  
plot(pop.infl, pch=19, xlab="Unit", ylab="Influence",   
 main="Effect on Standard Deviation")



This graph shows that values become more influential as they get closer to either extreme. On the other hand, values closer to the mean have negative influence, meaning they work to reduce the standard deviation.