Week 2 Excercise

Jacques Uber

4/9/2020

Here is the solution to 3.40.

Define the data

First we load 2010 U.S. Census data for a random sample of 40 households where each value represents the age of the first person recorded for that household.

```r
data <- c(
  42, 29, 31, 38, 55, 27, 28,
  33, 49, 70, 25, 21, 38, 47,
  63, 22, 38, 52, 50, 41, 19,
  22, 29, 81, 52, 26, 35, 38,
  29, 31, 48, 26, 33, 42, 58,
  40, 32, 24, 34, 25
)
```

The problem asks us to calculate P_{30}, P_{80}, Q_1, Q_3, the range, and the interquartile range for our data.

Define a helper function

R provides a `quantile(..., type=2)` function that can easily help us calculate the values we need, but for the sake of my own learning, lets define our own function using the definition provided in Black on page 50.

```r
myQuartile = function (xs, P) {
  # xs is assumed to be a vector
  # P is the percentile we are trying to fine
  xs = sort(xs)
  i = (P / 100) * length(xs)
  if (i %% 1 == 0) {
    # i is a whole number
    if (i + 1 >= length(xs)) {
      result = xs[i]
    } else {
      result = (xs[i] + xs[i + 1]) / 2
    }
  } else {
    # i is not a whole number
    result = xs[floor(i + 1)]
  }

  return(result)
}
```

Solution

$P_{30} = 29$ (check answer 29)

$P_{80} = 49.5$ (check answer 49.5)

$Q_1 = 27.5$ (check answer 27.5)

$Q_3 = 47.5$ (check answer 47.5)

$range = 62$

$IQR = Q3 - Q1 = 20$

The code used to calculate this was:

P_{30} = myQuartile(data, 30) (check answer quantile(data, 0.30, type=2))

P_{80} = myQuartile(data, 80) (check answer quantile(data, 0.80, type=2))

Q_{1} = myQuartile(data, 25) (check answer quantile(data, 0.25, type=2))

Q_{3} = myQuartile(data, 75) (check answer quantile(data, 0.75, type=2))

$range$ = max(data) - min(data)

IQR = $Q{3} - Q{1}$ = r myQuartile(data, 75) - myQuartile(data, 25)