How to normalize data with Z-scores

David Montwé*[1]

[1]Department of Forest Resources Management, Faculty of Forestry, The University of British Columbia, Forest Sciences Centre, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

0.1 Equation to calculate z-scores

$Z_i = \frac{data_i - \bar{x}}{\sigma}$

where:

Z_i is the Z-score for value i; $value_i$ is the value for position i in vector *data*; $\bar{x}$ is the arithmetic mean of *data*, and σ is the standard deviation of *data*.

0.2 Example in R based on the `iris` data set.

```
#Load example data
data_iris = iris$Sepal.Length[iris$Species=='virginica']

# A vector with 50 numeric values
str(data_iris)
```

```
##  num [1:50] 6.3 5.8 7.1 6.3 6.5 7.6 4.9 7.3 6.7 7.2 ...
```

```
# Manual calculation of z-scores (see equation above)
Z_scores_m = (data_iris - mean(data_iris))/sd(data_iris)
str(Z_scores_m)
```

```
##  num [1:50] -0.453 -1.239 0.805 -0.453 -0.138 ...
```

```
# Using the scale() function - default arguments are:  center = TRUE, scale = TRUE
Z_scores_sc = scale(data_iris)
str(Z_scores_m)
```

```
##  num [1:50] -0.453 -1.239 0.805 -0.453 -0.138 ...
```

```
#Difference of 0
mean(Z_scores_m -Z_scores_sc)
```

```
## [1] 0
```