Z-scores: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2020

Syntax

- Writing a function that converts a value to a z-score:

```
standard_deviation <- function(vector) {

    distances  <-  (vector - mean(vector))**2

    sqrt(sum(distances) / length(distances) )

}

z_score <- function(value, vector, bessel = FALSE) {

    mean  <-  mean(vector)

    st_dev <- ifelse(!bessel, sd(vector), standard_deviation(vector))

    (value - mean) / st_dev

}
```

- Standardizing a varaiable in a dataframe:

```
df <- df %>%

    mutate(z_var = (var - mean(var)) / standard_deviation(var) )
```

- Transforming a standardized distribution to a different distribution, with a predefined mean and standard deviation:

```
mean  <-  some_mean

st_dev  <-  some_standard_deviation

df <- df %>%

        mutate(transformed_var = z_var*st_dev + mean )
```

Concepts

- A **z-score** is a number that describes the location of a value within a distribution. Non-zero z-scores (+1, -1.5, +2, -2, etc.) consist of two parts:
    - *A sign*, which indicates whether the value is above or below the mean.
    - *A value*, which indicates the number of standard deviations that a value is away from the mean.

- The z-score of the mean is 0.
- To compute the z-score for a value coming from a population with mean and standard deviation , we can use this formula:

- To compute the z-score for a value coming from a sample with mean and standard deviation , we can use this formula:

- We can **standardize** any distribution by transforming all its values to z-scores. The resulting distribution will have a mean of 0 and a standard deviation of 1. Standardized distributions are often called **standard distributions**.
- Standardization is useful for **comparing values** coming from distributions with different means and standard deviations.
- We can transform any population of z-scores with mean and to a distribution with any mean and any standard deviation by converting each z-score to a value using this formula:

- We can transform any sample of z-scores with mean and to a distribution with any mean and any standard deviation by converting each z-score to a value using this formula:

Resources

- Z-score calculation in R.

- The Wikipedia entry on z-scores.

Takeaways by Dataquest Labs, Inc. - All rights reserved © 2020