

Z-scores: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2021

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 variable 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 z for a value x coming from a population with mean μ and standard deviation σ , we can use this formula:

$$z = \frac{x - \mu}{\sigma}$$

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

$$z = \frac{x - \bar{x}}{s}$$

- 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 $\mu_z = 0$ and $\sigma_z = 1$ to a distribution with any mean μ and any standard deviation σ by converting each z-score z to a value x using this formula:

$$x = z\sigma + \mu$$

- We can transform any sample of z-scores with mean $\bar{x}_z = 0$ and $s_z = 1$ to a distribution with any mean \bar{x} and any standard deviation s by converting each z-score z to a value x using this formula:

$$x = zs + \bar{x}$$

Resources

- [Z-score calculation in R](#).
- [The Wikipedia entry on z-scores](#).