

## Project 8

# Calculate Z-scores

In statistics, the z-score (or standard score) of an observation is the number of standard deviations that it is above or below the population mean.

To calculate a z-score you must know the population mean and the population standard deviation. In cases where it is impossible to measure every observation of a population, you can estimate the standard deviation using a random sample.

Create a z-score visualization to answer questions like the following:

- What percentage of values fall below a specific value?
- What values can be considered exceptional? For example, in an IQ test, what scores represent the top five percent?
- What is the relative score of one distribution versus another? For example, Michael is taller than the average male and Emily is taller than the average female, but who is relatively taller within their gender?

As a general rule, z-scores lower than -1.96 or higher than 1.96 are considered unusual and interesting. That is, they are statistically significant outliers.

This article demonstrates how to calculate a z-score in Tableau.

1. Connect to the **Sample - Superstore** data source provided with Tableau Desktop.
2. Create a calculated field to calculate average sales.  
Choose **Analysis > Create Calculated Field** to open the calculation editor. Name the calculation **Average Sales** and type or paste the following in the formula area:  
`WINDOW_AVG(SUM([Sales]))`
3. Create another calculated field to calculate the standard deviation. Name the calculation **STDEVP Sales** and type or paste the following in the

formula area:

```
WINDOW_STDEVP(SUM([Sales]))
```

4. Create one more calculated field, this one to calculate the z-score. Name the calculation **Z-score** and type or paste the following in the formula area:

```
(SUM([Sales]) - [Average Sales]) / [STDEVP Sales]
```

5. Drag **Z-Score** from the **Data** pane to **Columns** and **State** to **Rows**. Notice that the **Z-score** field on Columns has a table calculation icon on the right side (that is, a small triangle):



The **STDEVP Sales** function is based on the **WINDOW\_STDEVP** function, which is a table calculation function. The **Z-Score** function, in turn, is a table calculation function because it includes **STDEVP Sales** in its definition. When you use a calculated field that includes a table calculation function in a view, it's the same as adding a table calculation to a field manually. You can edit the field as a table calculation. In fact, that's what you do next.

6. Click the **Z-score** field on Columns and choose **Compute Using > State**. This causes the z-scores to be computed on a per-state basis.
7. Click the **Sort Descending** icon on the toolbar:



8. Hold down the Ctrl key and drag the **Z-score** field from **Columns** to **Color**.  
Ctrl + Drag copies a field as currently configured to an additional location.
9. Ctrl + Drag **Z-score** from Columns once again. This time drop it on **Label**.

You now have a distribution of z-scores broken out by state. California and New York both have z-scores greater than 1.96. You could conclude from this that California and New York have significantly higher average sales than other states.

