Z-scores: Takeaways

Syntax

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

```
def z_score(value, array, bessel = 0):
    mean = sum(array) / len(array)

    from numpy import std
    st_dev = std(array, ddof = bessel)

    distance = value - mean
    z = distance / st_dev

    return z
```

- Standardizing a `Series`:

```
standardized_distro = Series.apply(
                         lambda x: (x - Series.mean()) /
Series.std()
                         )
```

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

```
mean = some_mean
st_dev = some_standard_deviation

standardized_distro = Series.apply(
                         lambda z: z * 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 $\overline{x}$ and standard deviation s, we can use this formula:

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

- 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 $\overline{x}_z = 0$ and $s_z = 1$ to a distribution with any mean $\overline{x}$ and any standard deviation s by converting each z-score z to a value x using this formula:

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

Resources

- The `z-score()` function from `scipy.stats.mstats` — useful for standardizing distributions.

- The Wikipedia entry on z-scores.