

# نرمال سازی و استاندارد سازی داده ها در آمار

علیرضا سلطانی نشان

۱۱ آبان ۱۴۰۳

هیچ وقت ما نمی توانیم به صورت پیش فرض دو مقداری که دارای معیارهای مختلفی هستند را با یکدیگر محاسبه کنیم. برای مثال هیچ وقت نمی توانیم حاصل ضرب زیر را بدست آوریم:

$$40_{km/h} \times 100_{kg} \quad (۱)$$

زیرا وقتی دو ویژگی مانند سرعت (کیلومتر در ساعت) و وزن (کیلوگرم) را باهم مقایسه یا ترکیب می کنیم، تنها نرمالیزه کردن کافی نمی باشد و ممکن است به یک مفهوم فیزیکی نرسیم. برای چنین شرایطی که مقیاس ها و واحدها متفاوت هستند، چند راهکار وجود دارد که می تواند به استاندارد سازی کمک کند:

## بی واحد کردن داده ها یا Dimensionless Scaling

برای ایجاد مقیاس های یکسان، می توانیم داده ها را با استفاده از میانگین یا مقادیر ماکسیمم و مینیمم خودشان بی واحد کنیم.

$$\frac{(\frac{speed}{speed_{max}}) + (\frac{weight}{weight_{max}})}{2} \quad (۲)$$

$$\frac{(\frac{40}{80_{max}}) + (\frac{100}{120_{max}})}{2} = 0.6665 \quad (۳)$$

بر اساس معادله بالا داده های ما مقیاسی بدون واحد دارند. در نظر داشته باشیم که با استفاده از محدوده داده های فراوان می توانیم به روش های مختلفی محاسبات خود را انجام دهیم [۱].

- سرعت:  $40_{km/h}$
- وزن:  $100_{kg}$
- فرض کنیم که از همان مثال بالا در دیتاست خود اطلاعات زیر را هم داریم:
- میانگین سرعت در داده ها:  $50_{km/h}$
- انحراف معیار سرعت:  $20_{km/h}$
- میانگین وزن:  $80_{kg}$
- انحراف معیار وزن:  $25_{kg}$

## ۱ استانداردسازی با روش Z-score

هر مقدار را به صورت تفاضل از میانگین و تقسیم بر انحراف معیار، نرمال سازی می کنیم تا نتیجه Z-score بدست آید:

$$\frac{x - \mu}{\sigma} \quad (۴)$$

برای سرعت  $40_{km/h}$  خواهیم داشت:

$$\frac{40 - 50}{20} = -0.5 \quad (۵)$$

برای وزن  $100_{kg}$  خواهیم داشت:

$$\frac{100 - 80}{25} = 0.8 \quad (۶)$$

حال دو مقداری که از روش نرمال سازی Z-score حاصل شده است را می توان به عنوان مقادیر بدون واحد در نظر گرفت و تمام مقایسه ها و تحلیل های مورد نظر خود را روی آن ها انجام داد.

## ۲ واحدهای نسبی یا Relative units

در واحدهای نسبی ما باید مقدار ورودی مسئله را با یک مقدار استاندارد یا مبنا در یک جامعه خاص بسنجیم:

- مقدار استاندارد سرعت به عنوان ورودی  $40_{km/h}$  بوده است و در جامعه محدوده این داده  $80_{km/h}$  می باشد.
  - همچنین در محاسبه وزن داده ورود  $100_{kg}$  بوده است و در جامعه محدوده این داده  $120_{kg}$  می باشد.
- فرمول:

$$\frac{speed}{basespeed} = \frac{40}{80} = 0.5 \quad (۷)$$

$$\frac{weight}{baseweight} = \frac{100}{120} = 0.833 \quad (۸)$$

حال نتایج بدست آمده کاملاً بدون واحد هستند و با روش واحدهای نسبی محاسبه شده اند:

- سرعت نسبی:  $0.5_{km/h}$
- وزن نسبی:  $0.833_{kg}$

## مراجع

- [1] Wikipedia contributors. Normalization (statistics) — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Normalization\\_\(statistics\)&oldid=1203519063](https://en.wikipedia.org/w/index.php?title=Normalization_(statistics)&oldid=1203519063), 2024. [Online; accessed 1-November-2024].