Measures of asymmetry (skewness)

Formula

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}^{\,3}}$$

Formula in terms of moments

Let μ_r be the r-th central moment, defined as:

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^r$$

> Note: central moment is the one where a = $\bar{x}$

For skewness:

- Moment-based (population) skewness:

$$\frac{\mu_3}{\mu_2^{3/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{3/2}}$$

- Sample skewness (with Bessel correction in the denominator):

$$\frac{\mu_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{3/2}}$$

Meaning of skewness

Skewness indicates whether the data is concentrated on one side (has a tail).

- When mean is bigger than a median, we say that we have a positive (right) skew. In those cases on the graphs we can tell that the tail is leaning towards the right side.

- When mean = median = mode, there's no skew. We say that the distribution is symmetrical.
- As the mean is lower than the meadian we say that there's a negative or left skew. The outliers are to the left.

Why skew is important

- Skewness tells us a lot about where data is situated and cluttered the most.
- It highlights the presence and impact of outliers.
- Measures of asymmetry like skewness are the link between central tendency measures and probability theory.