Statistics

Monday, February 10, 2025 10:22 AM

Descriptive

Mean – Not robust to outliers at all

Calculate proportions as an alternative for categorical variables.

Weighted Mean : $\dfrac{\Sigma x_i w_i}{\Sigma w_i}$

Geometric mean : $\sqrt[n]{x_1 x_2 x_3 \cdots x_n}$ → Used in avg. rates of return

Harmonic mean : $\dfrac{n}{x_1 + x_2 + \cdots x_n}$

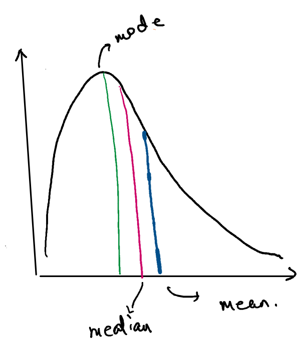
Median – Much more robust.

$[1, 2, 3, 4, 5, 6]$.

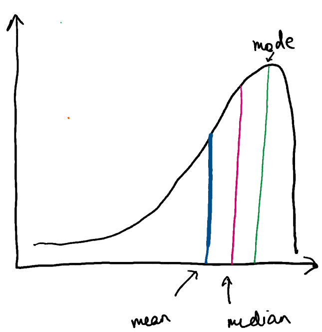
$\dfrac{3+4}{2} = 3.5$

Mode: It is basically the most fashionable observation.

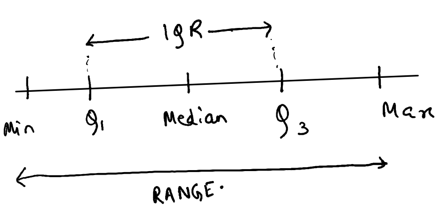
→ It is horrible for small samples.



Right / +ve



Left / -ve.



Variance

How the data varies from its mean :

data point :

How the data varies from its mean:

$$Var = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

[When there is 1 data point: no variance, 2 data points: $\frac{(x_i - \bar{x})^2}{1}$.
$\therefore$ n data points: $\frac{\Sigma (x_i - \bar{x})^2}{n-1}$]

$$Std. Dev = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$$

When data is varied,
[1, 2, 3] vs [101, 102, 103],
the value for $x_i - \bar{x}$ remains same in both cases (just scaled)
However, the spreads are not exactly the same.
Coeff. of Variation helps capture that.

$$CV = \frac{std.\ dev}{\bar{x}}$$

Skewness

1) $Skew = \frac{3(mean - median)}{std.\ dev} = \frac{mean - mode}{std.\ dev}$

mode = 3 median − 2 mean.

[3 (mean − median) = mean − mode.
$\Rightarrow$ 3 mean − 3 median = mean − mode
$\Rightarrow$ 2 mean = 3 median − mode.
$\Rightarrow$ mode = 3 median − 2 mean.]

$$Population\ Skew = \frac{1}{n} \frac{\Sigma (x - \mu)^3}{\sigma^3}$$

$$sample\ skew = \frac{n}{(n-1)(n-2)} \frac{\Sigma (x - \bar{x})^3}{s^3}$$

Kurtosis (How peaked?).

$$Pop.\ Kurtosis = \frac{1}{n} \frac{\Sigma (x - \mu)^4}{\sigma^4}$$

$$Sample\ Kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \left(\frac{\Sigma (x - \bar{x})^4}{s^4} \right) - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$(n-1)(n-2)(n-3)$ \ s^4 / $(n-2)(\ldots)$

$K = 3$ meso

$K > 3$ lepto

$K < 3$ plato

Standard Error

It is the standard deviation of means.

$$SE = \frac{\text{std. dev}}{\sqrt{n}}$$

$$SE = \sqrt{\frac{p(1-p)}{n}}$$