

Regola empirica sulla concentrazione dati

Quando l'istogramma dei dati è ragionevolmente normale (assomiglia ad una gaussiana), i dati si trovano:

- per il 68% entro σ dalla media;
- per il 95% entro 2σ ;
- per il 99.7% entro 3σ .

Infatti:

- se $X \sim N(0, 1)$, $P(-1 \leq X \leq 1) \simeq 0.68$;
- se $Y \sim N(m, \sigma^2)$,

$$P(m - \sigma \leq Y \leq m + \sigma) = P\left(-1 \leq \underbrace{\frac{y - m}{\sigma}}_X \leq 1\right) = 0.68.$$