

# Métodos para Particionar Datos

## 1 Regla de Sturges

$$k = 1 + \log_2(n) \quad (1)$$

Es simple y funciona bien para conjuntos de datos pequeños y moderados. Asume una distribución normal de los datos.

Desventajas: Tiende a subestimar el número de clases para conjuntos de datos grandes.

## 2 Regla de Scott

$$h = 3.5\sigma n^{-\frac{1}{3}} \quad (2)$$

Minimiza la varianza de la estimación del histograma. Es útil para distribuciones aproximadamente normales.

Desventajas: Puede no funcionar bien con datos que tienen una distribución sesgada.

## 3 Regla de Freedman-Diaconis

$$h = 2 \times \frac{IQR}{n^{\frac{1}{3}}} \quad (3)$$

Utiliza el rango intercuartílico (IQR) para reducir el impacto de los valores atípicos, siendo más robusto que la regla de Scott.

Desventajas: Puede resultar en un número excesivo de intervalos si el conjunto de datos es pequeño.

## 4 Regla de Doane

$$k = 1 + \log_2(n) + \log_2 \left( 1 + \frac{|g_1|}{\sqrt{\frac{6(n-2)}{(n+1)(n+3)}}} \right) \quad (4)$$

Es una extensión de la regla de Sturges que corrige para el sesgo en la distribución de los datos. Es útil para distribuciones no normales.

Desventajas: A veces puede sobreestimar el número de clases.

## 5 Regla de Rice

$$k = 2 \times n^{\frac{1}{3}} \quad (5)$$

Es una regla empírica simple y más conservadora que la de Sturges. Se usa a menudo como una alternativa rápida y simple.

Desventajas: Tiende a sobreestimar el número de intervalos para conjuntos de datos pequeños.

## 6 Conclusión

La elección del método depende de la naturaleza de los datos. Para datos con distribución normal, la regla de Scott o Sturges podría ser suficiente. Para datos con valores atípicos o distribuciones sesgadas, Freedman-Diaconis es más robusta. En general, es recomendable probar diferentes métodos y elegir el que mejor refleje la distribución de los datos.