

INTRO TO STATISTICS

la moda se suele ver con valores categóricos

Cuando la distribución de los datos es simétrica, USAR MEDIA

↓
Cuando no lo es, USAR MEDIANA

Medidas de dispersión → Varianza ($A^2 \rightarrow A$ desviación)

Desviación estándar

`np.std(ddof=1)`

Percentiles

`np.quantile(datos, 0.5)`

Rango intercuartílico

`scipy.stats.iqr()`

Es la distancia entre los percentiles 25 y 75

Outliers, se consideran valores que

$data < (Q_1 - 1.5 * IQR)$

$data > (Q_3 + 1.5 * IQR)$

`df.sample(n)` → N valores.
→ Elige un valor aleatorio

Distribuciones continuas

Dist. uniforme

→ Para ver la probabilidad de 7 o menos

`scipy.stats.uniform.cdf(7, 0, 10)`

100 valores

límite superior

la distribución binomial

Para generar valores aleatorios 0 ó 1

`scipy.stats.binom.rvs(1, 0.5, size=N)`

si no quieres 50/50 póngale otro

la distribución normal

`scipy.stats.norm.cdf(N*interos, media, desviacion estandar)`

Probabilidad de por debajo del N

`scipy.stats.norm.ppf(Probabil, media, desv. std)`

`scipy.stats.norm.rvs(media, desv. std, size=N)`

→ Generación de valores aleatorios

la distribución de poisson

`poisson.pmf(valor, media)`

`poisson.cdf(valor o mas, media)`

`poisson.rvs(media, size=N)`

`expon.cdf(valor, scale = 1/x)`

Otras distribuciones

→ exponencial

→ t-distribution

→ Log-normal

INTRO TO STATISTICS

Correlation

Pour voir la corrélation entre 2 variables →

$$\text{df}[\text{col.1}] \cdot \text{corr}(\text{df}[\text{col.2}])$$

→ Valeur du coefficient r