



International Center for Tropical Agriculture
Since 1967 *Science to cultivate change*

Análisis estadístico descriptivo

Lima, Perú

Octubre 2019

Autores:

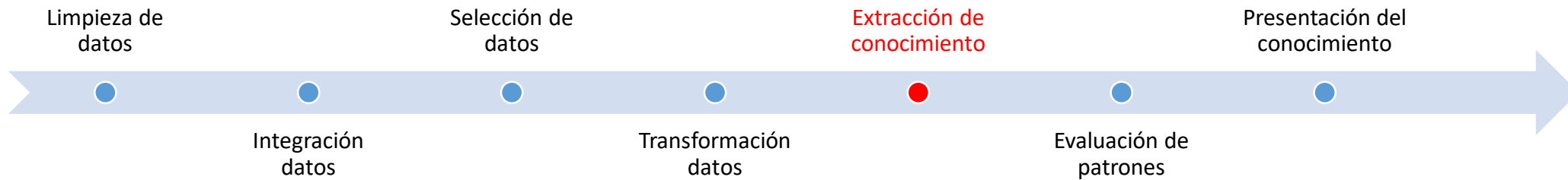
Hugo Andrés Dorado, Juan Camilo Rivera

h.a.dorado@cgiar.org , j.c.rivera@cgiar.org



Estadística frente a la minería de datos

Minería de datos: Extracción de conocimiento de grandes cantidades de datos. *Han y Kamber (2001)*.



Estadística: se ha orientado a la verificación y validación de hipótesis.

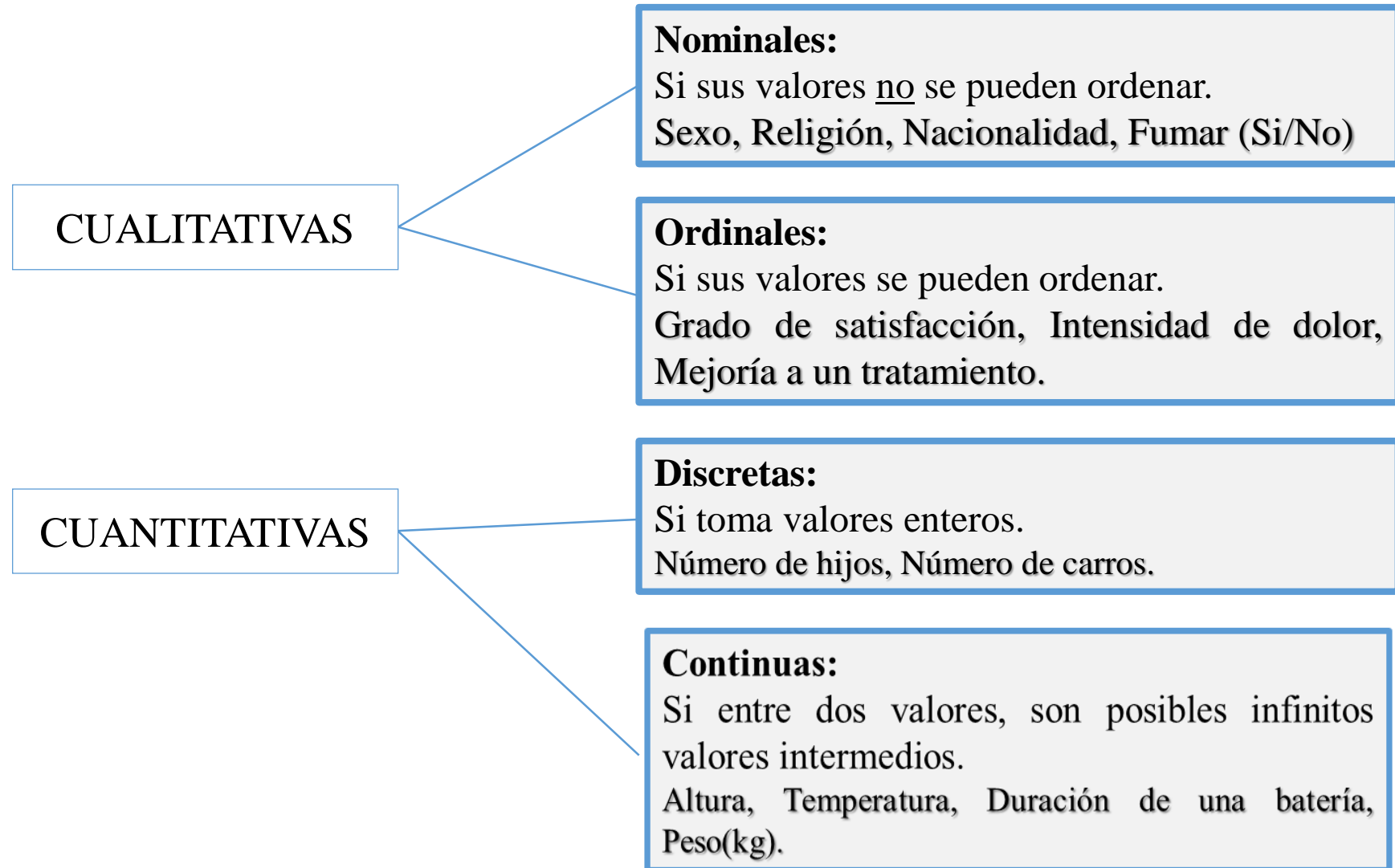
Estadística descriptiva

Distribución de las variables

Medidas de tendencia central y
dispersión

Visualizar gráficamente

Tipo de variables



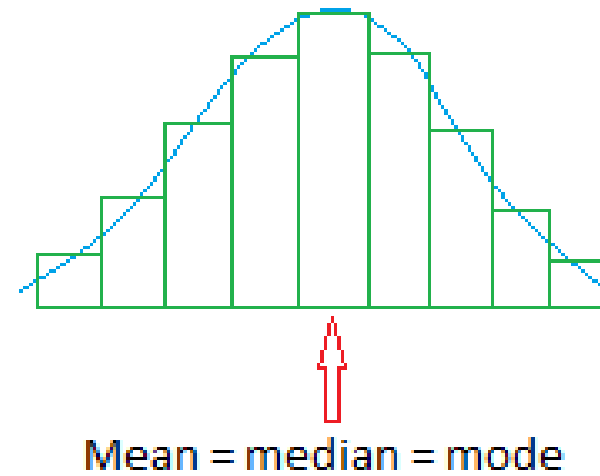
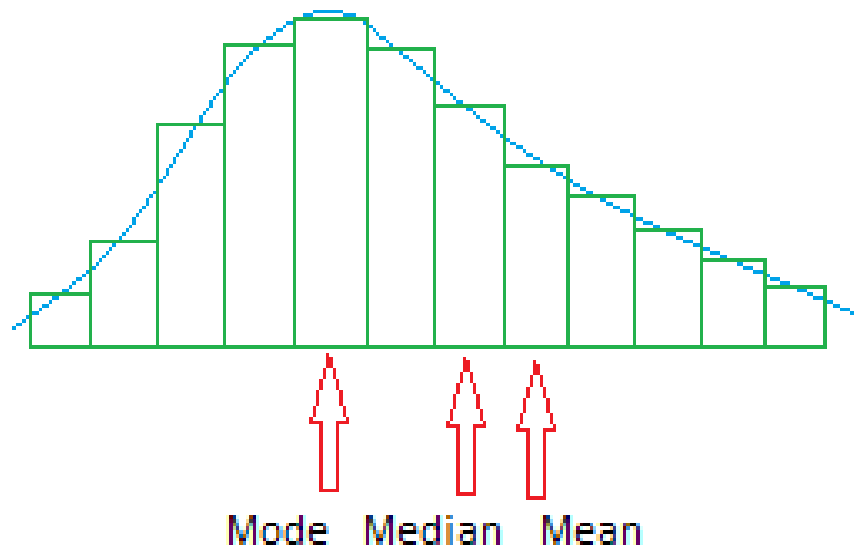
El código en R

MoraEspinas	Departamento	Ciudad	Elevacion	Pendiente	SR_1
Min. :0.000	Ica :329	Atenas : 82	Min. : 0.00	Min. : 2.00	9,17 : 6
1st Qu.:1.000	Loreto : 62	Iquitos: 41	1st Qu.: 28.00	1st Qu.:11.00	10,16 : 5
Median :1.000	I cca : 52	Nauta : 40	Median : 60.00	Median :15.00	11,43 : 5
Mean :0.793	I k : 17	Patras :325	Mean : 57.62	Mean :20.02	11,51 : 5
3rd Qu.:1.000	I ca : 9		3rd Qu.: 84.00	3rd Qu.:33.00	4,91 : 5
Max. :1.000	L ore to: 8		Max. :119.00	Max. :50.00	9,12 : 5
	(other) : 11				(other):457

MoraEspinas	Departamento	Ciudad	Elevacion	Pendiente	SR_1
Espinas :101	Ica :407	Atenas : 82	Min. : 0.00	Min. : 2.00	Min. : 4.000
NoEspinas:387	Loreto: 81	Iquitos: 41	1st Qu.: 28.00	1st Qu.:11.00	1st Qu.: 9.268
		Nauta : 40	Median : 60.00	Median :15.00	Median :10.120
		Patras :325	Mean : 57.62	Mean :20.02	Mean : 9.475
			3rd Qu.: 84.00	3rd Qu.:33.00	3rd Qu.:11.102
			Max. :119.00	Max. :50.00	Max. :12.000

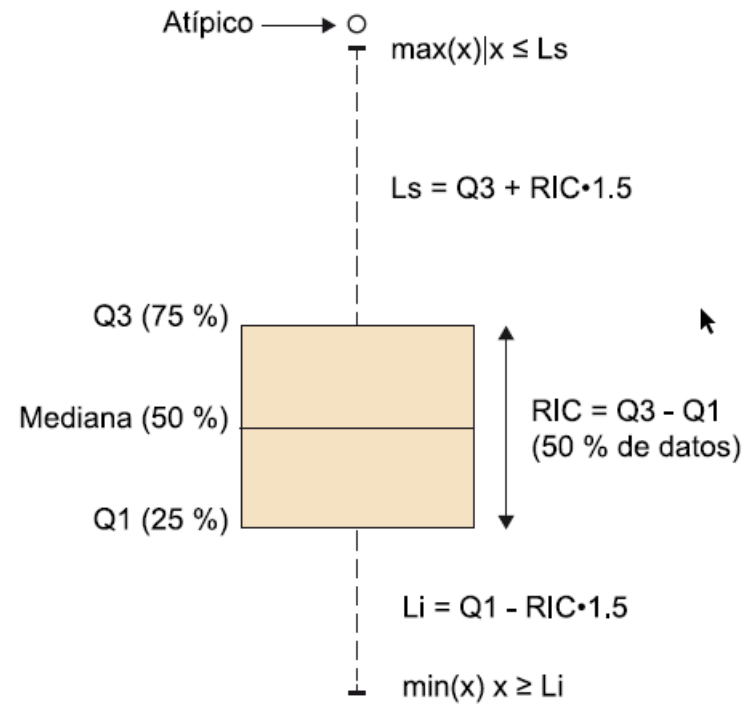
Estimadores

Tendencia central



Estimadores

Boxplot



Estimadores

Tendencia central y dispersión robustos

Medidas	Explicación	Código en R
Mediana	La mitad de una lista ordenada de números.	median()
Media recortada	Eliminar una cantidad de los valores extremos.	mean()
Media winsorizada	Los valores extremos son reemplazados por el valor menor o mayor que ha quedado.	mean(trim=0.05)
Rango intercuartílico	La diferencia entre el primer y tercer cuantil.	IQR()
Desviación Estándar	Mide dispersion de un conjunto de valores.	sd()

El código en R

- Pegar aca el summary de histograma