Medidas estadísticas

Medida	*Poblacional*	*Muestral*
	$Población\ ordenada = (x_1, x_2, .., x_N)$	$Muestra\ ordenada = (X_1, X_2, .., X_n)$
Tamaño (número de elementos)	N	n
Media (aritmética)	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
Moda	Valor x_i que más se repite	Valor X_i que más se repite
Cuantil p	Si $p \cdot N$ no es un número entero: $q_p = x_{\lfloor p \cdot N \rfloor + 1}$ Si $p \cdot N$ es un número entero: $q_p = \frac{x_{p \cdot N} + x_{p \cdot N+1}}{2}$ $Donde$ $\lfloor p \cdot N \rfloor =$ parte entera de $p \cdot N$	Si $p \cdot n$ no es un número entero: $\hat{q}_p = x_{\lfloor p \cdot n \rfloor + 1}$ Si $p \cdot n$ es un número entero: $\hat{q}_p = \frac{X_{p \cdot n} + X_{p \cdot n+1}}{2}$ $Donde$ $\lfloor p \cdot n \rfloor =$ parte entera de $p \cdot n$
Valor mínimo, valor máximo, rango	$\min x = x_1, \max x = x_N$ $\text{rango } x = x_N - x_1$	$\min X = X_1, \max X = X_n$ $\text{rango } X = X_n - X_1$
Varianza	$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	También llamada cuasivarianza muestral $s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$
Desviación estándar	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$
Coeficiente de variación	$cv = \frac{\sigma}{\mu}$	$CV = \frac{s}{\bar{X}}$
Rango intercuartílico	$IQR = Q_3 - Q_1$ Regla datos atípicos: $[Q_1 - 1.5 \cdot IQR,\ Q_3 + 1.5 \cdot IQR]$	$\widehat{IQR} = \hat{Q}_3 - \hat{Q}_1$ Regla datos atípicos: $\left[\hat{Q}_1 - 1.5 \cdot \widehat{IQR},\ \hat{Q}_3 + 1.5 \cdot \widehat{IQR}\right]$
Sesgo o coeficiente de asimetría	$A = \frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^3}{\sigma^3}$	$\hat{A} = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^3}{s^3}$
Curtosis o coeficiente de apuntamiento	$K = \frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^4}{\sigma^4} - 3$	$\hat{K} = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^4}{s^4} - 3$
Proporción de elementos que cumplen una condición	$p = \frac{n^{\underline{o}}\ elementos\ que\ cumplen\ la\ condición}{N}$	$\hat{p} = \frac{n^{\underline{o}}\ elementos\ que\ cumplen\ la\ condición}{n}$

Medidas estadísticas a partir de una tabla de frecuencias

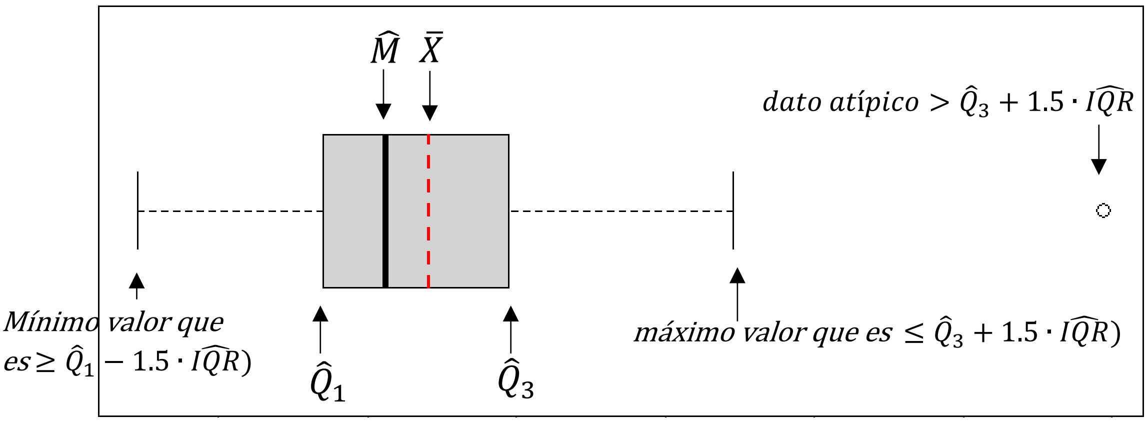
$Fila$	x ó X	f	h	F	H
1	v_1	f_1	h_1	F_1	H_1
2	v_2	f_2	h_2	F_2	H_2
...	...	...	...	...	...
i	v_i	f_i	h_i	F_i	H_i
...	...	...	...	...	...
k	v_k	f_k	h_k	F_k	H_k

$Fila$	x ó X	$Marca$	f	h	F	H
1	$[a_1, b_1)$	m_1	f_1	h_1	F_1	H_1
2	$[a_2, b_2)$	m_2	f_2	h_2	F_2	H_2
...	...	...	...	...	...	...
i	$[a_i, b_i)$	m_i	f_i	h_i	F_i	H_i
...	...	...	...	...	...	...
k	$[a_k, b_k]$	m_k	f_k	h_k	F_k	H_k

Regla de Sturges: $k = \lceil 1 + 3{,}322 \log_{10} N \rceil$ (población) ó $k = \lceil 1 + 3{,}322 \log_{10} n \rceil$ (muestra)
Ejemplos: $10 \le n \le 15 \ \rightarrow k = 5$; $16 \le n \le 31 \ \rightarrow k = 6$; $32 \le n \le 63 \ \rightarrow k = 7$; $64 \le n \le 127 \ \rightarrow k = 8$

Medida	*Poblacional*	*Muestral*
Media (aritmética)	$\mu = \frac{\sum_{i=1}^{k} f_i v_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{k} f_i v_i}{n}$
Media (aritmética) con datos agrupados	$\mu = \frac{\sum_{i=1}^{k} f_i m_i}{N}$ $m_i = \frac{a_i + b_i}{2}$	$\bar{X} = \frac{\sum_{i=1}^{k} f_i m_i}{n}$
Varianza	$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (v_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{k} f_i (v_i - \bar{X})^2}{n-1}$
Varianza con datos agrupados	$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \bar{X})^2}{n-1}$
Cuantil	Siendo v_i el primer valor que cumple: $F_i \ge p \cdot N$ Si $F_i > p \cdot N$: $q_p = v_i$ Si $F_i = p \cdot N$: $q_p = \frac{v_i + v_{i+1}}{2}$	Siendo v_i el primer valor que: $F_i \ge p \cdot n$ Si $F_i > p \cdot N$: $\hat{q}_p = v_i$ Si $F_i = p \cdot N$: $\hat{q}_p = \frac{v_i + v_{i+1}}{2}$
Cuantil con datos agrupados	Siendo $[a_i, b_i)$ el primero que cumple: $F_i \ge p \cdot N$ $q_p = a_i + (b_i - a_i) \cdot \frac{p \cdot N - F_{i-1}}{f_i}$	Siendo $[a_i, b_i)$ el primero que: $F_i \ge p \cdot n$ $\hat{q}_p = a_i + (b_i - a_i) \cdot \frac{p \cdot n - F_{i-1}}{f_i}$

Diagrama de caja para una muestra



Estadística bidimensional

Medida	***Poblacional***	***Muestral***
Tamaño	N $Población = \left((x_1, y_1), (x_2, y_2), .., (x_N, y_N)\right)$	n $Muestra = \left((X_1, Y_1), (X_2, Y_2), .., (X_n, Y_n)\right)$
Medias (aritméticas) marginales	$\mu_x = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
	$\mu_y = \frac{\sum_{i=1}^{N} y_i}{N}$	$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$
Varianzas marginales	${\sigma_x}^2 = \frac{\sum_{i=1}^{N}(x_i - \mu_x)^2}{N}$	${s_X}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$
	${\sigma_y}^2 = \frac{\sum_{i=1}^{N}\left(y_i - \mu_y\right)^2}{N}$	${s_Y}^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$
Covarianza	$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)\left(y_i - \mu_y\right)}{N}$	$s_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$
Coeficiente de correlación (de Pearson)	$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$r = \frac{s_{XY}}{s_X s_Y}$

Medidas estadísticas a partir de una tabla de contingencia sin datos agrupados

	w_1	w_2	…	w_j	…	w_l	f_X	h_X
v_1	$f_{1,1}$	$f_{1,2}$	…	$f_{1,j}$	…	$f_{1,l}$	f_{v_1}	h_{v_1}
v_2	$f_{2,1}$	$f_{2,2}$	…	$f_{2,j}$	…	$f_{2,l}$	f_{v_2}	h_{v_2}
…	…	…	…	…	…	…	…	…
v_i	$f_{i,1}$	$f_{i,2}$	…	$f_{i,j}$	…	$f_{i,l}$	f_{v_i}	h_{v_i}
…	…	…	…	…	…	…	…	…
v_k	$f_{k,1}$	$f_{k,2}$	…	$f_{k,j}$	…	$f_{k,l}$	f_{v_k}	h_v
f_Y	f_{w_1}	f_{w_2}	…	f_{w_j}	…	f_{w_l}	N ó n	
h_Y	h_{w_1}	h_{w_2}	…	h_{w_j}	…	h_{w_l}		1

Medida	***Poblacional***	***Muestral***
Media (aritmética) marginal de x (ó X)	$\mu_x = \frac{\sum_{i=1}^{k} f_{v_i} v_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{k} f_{v_i} v_i}{n}$
Varianza marginal de x (ó X)	${\sigma_x}^2 = \frac{\sum_{i=1}^{k} f_{v_i} (v_i - \mu_x)^2}{N}$	${s_X}^2 = \frac{\sum_{i=1}^{k} f_{v_i} (v_i - \bar{X})^2}{n-1}$
Cuantil marginal de x (ó X)	Siendo v_i el primer valor que: $F_{v_i} \geq p_x \cdot N$ Si $F_{v_i} > p_x \cdot N$: $q_{p_x} = v_i$ Si $F_{v_i} = p_x \cdot N$: $q_{p_x} = \frac{v_i + v_{i+1}}{2}$ Donde $F_{v_i} = \sum_{r=1}^{i} f_{v_i}$	Siendo v_i el primer valor que: $F_{v_i} \geq p_X \cdot n$ Si $F_{v_i} > p_X \cdot n$: $\hat{q}_{p_X} = v_i$ Si $F_{v_i} = p_X \cdot n$: $\hat{q}_{p_X} = \frac{v_i + v_{i+1}}{2}$ Donde $F_{v_i} = \sum_{r=1}^{i} f_{v_i}$
Media condicionada de x (ó X) cuando y (ó Y) = w_j	$\mu_x \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} v_i}{f_{w_j}}$	$\bar{X} \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} v_i}{f_{w_j}}$
Varianza condicionada de x (ó X) cuando y (ó Y) = w_j	${\sigma_x}^2 \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} \left(v_i - \mu_x \vert w_j\right)^2}{f_{w_j}}$	${s_X}^2 \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} \left(v_i - \bar{X} \vert w_j\right)^2}{f_{w_j} - 1}$
Cuantil condicionado p de x (ó X) cuando y (ó Y) = w_j	Siendo v_i el primer valor que: $F_{i,j} \geq p_x \cdot f_{w_j}$ Si $F_{i,j} > p_x \cdot f_{w_j}$: $q_{p_x} \vert w_j = v_i$ Si $F_{i,j} = p_x \cdot f_{w_j}$: $q_{p_x} \vert w_j = \frac{v_i + v_{i+1}}{2}$ Donde $F_{i,j} = \sum_{r=1}^{i} f_{r,j}$	Siendo v_i el primer valor que: $F_{i,j} \geq p_X \cdot f_{w_j}$ Si $F_{i,j} > p_X \cdot f_{w_j}$: $\hat{q}_{p_X} \vert w_j = v_i$ Si $F_{i,j} = p_X \cdot f_{w_j}$: $\hat{q}_{p_X} \vert w_j = \frac{v_i + v_{i+1}}{2}$ Donde $F_{i,j} = \sum_{r=1}^{i} f_{r,j}$
Independencia	$\frac{f_{i,j}}{N} = h_{v_i} \cdot h_{w_j}$ para todo i y j	$\frac{f_{i,j}}{n} = h_{v_i} \cdot h_{w_j}$ para todo i y j

Medidas estadísticas a partir de una tabla de contingencia con datos agrupados

	w_1	w_2	...	w_j	...	w_l	f_X	h_X
$c_1 = [a_1, b_1)$ m_1	$f_{1,1}$	$f_{1,2}$	...	$f_{1,j}$	...	$f_{1,l}$	f_{c_1}	h_{c_1}
$c_2 = [a_2, b_2)$ m_2	$f_{2,1}$	$f_{2,2}$	...	$f_{2,j}$	...	$f_{2,l}$	f_{c_2}	h_{c_2}
...	...	...	...	...	...	...	...	...
$c_i = [a_i, b_i)$ m_i	$f_{i,1}$	$f_{i,2}$	...	$f_{i,j}$	...	$f_{i,l}$	f_{c_i}	h_{c_i}
...	...	...	...	...	...	...	...	...
$c_k = [a_k, b_k]$ m_k	$f_{k,1}$	$f_{k,2}$	...	$f_{k,j}$	...	$f_{k,l}$	f_{c_k}	h_{c_k}
f_Y	f_{w_1}	f_{w_2}	...	f_{w_j}	...	f_{w_l}	N ó n	
h_Y	h_{Y_1}	h_{Y_2}	...	h_{Y_j}	...	h_{Y_l}		1

Medida	***Poblacional***	***Muestral***
Media (aritmética) marginal de x (ó X)	$\mu_x = \frac{\sum_{i=1}^{k} f_{c_i} m_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{k} f_{c_i} m_i}{n}$
Varianza marginal de x (ó X)	${\sigma_x}^2 = \frac{\sum_{i=1}^{k} f_{c_i} (m_i - \mu_x)^2}{N}$	${s_X}^2 = \frac{\sum_{i=1}^{k} f_{c_i} (m_i - \bar{X})^2}{n-1}$
Cuantil marginal p de x (ó X)	Siendo $[a_i, b_i)$ el primero que cumple: $F_{c_i} \geq p_x \cdot N$ $q_{p_x} = a_i + (b_i - a_i) \cdot \frac{p_x \cdot N - F_{c_{i-1}}}{f_{c_i}}$ Donde $F_{c_i} = \sum_{r=1}^{i} f_{c_i}$	Siendo $[a_i, b_i)$ el primero que cumple: $F_{c_i} \geq p_X \cdot n$ $\hat{q}_{p_X} = a_i + (b_i - a_i) \cdot \frac{p_X \cdot n - F_{c_{i-1}}}{f_{c_i}}$ Donde $F_{c_i} = \sum_{r=1}^{i} f_{c_i}$
Media condicionada de x (ó X) cuando y (ó Y) = w_j	$\mu_x \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} m_i}{f_{w_j}}$	$\bar{X} \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} m_i}{f_{w_j}}$
Varianza condicionada de x (ó X) cuando y (ó Y) = w_j	${\sigma_x}^2 \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} \left(m_i - \mu_x \vert w_j\right)^2}{f_{w_j}}$	${s_X}^2 \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} \left(m_i - \bar{X} \vert w_j\right)^2}{f_{w_j} - 1}$
Cuantil condicionado p de x (ó X) cuando y (ó Y) = w_j	Siendo $[a_i, b_i)$ el primero que cumple: $F_{i.j} \geq p_x \cdot f_{w_j}$ $q_{p_x} \vert w_j = a_i + (b_i - a_i) \cdot \frac{p_x \cdot f_{w_j} - F_{i-1,j}}{f_{i,j}}$ Donde $F_{i,j} = \sum_{r=1}^{i} f_{r,j}$	Siendo $[a_i, b_i)$ el primero que cumple: $F_{i,j} \geq p_X \cdot f_{w_j}$ $\hat{q}_{p_X} \vert w_j = a_i + (b_i - a_i) \cdot \frac{p_X \cdot f_{w_j} - F_{i-1,j}}{f_{i,j}}$ Donde $F_{i,j} = \sum_{r=1}^{i} f_{r,j}$
Independencia	$\frac{f_{i,j}}{N} = h_{c_i} \cdot h_{w_j}$ para todo i y j	$\frac{f_{i,j}}{n} = h_{c_i} \cdot h_{w_j}$ para todo i y j