Medidas estadísticas

Medida	*Poblacional*	*Muestral*
Tamaño (número de elementos)	N $Población = (x_1, x_2, .., x_N)$	n $Muestra = (X_1, X_2, .., X_n)$
Medidas de centralización		
Media (aritmética)	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
Mediana	$Población\ ordenada = (x_1, x_2, .., x_N)$ Si $N/2$ no es un número entero: $M = x_{\lfloor N/2 \rfloor + 1}$ Si $N/2$ es un número entero: $M = \frac{x_{N/2} + x_{N/2+1}}{2}$	$Muestra\ ordenada = (X_1, X_2, .., X_n)$ Si n/2 no es un número entero: $\widehat{M} = X_{\lfloor n/2 \rfloor + 1}$ Si n/2 es un número entero: $\widehat{M} = \frac{X_{n/2} + X_{n/2+1}}{2}$
Moda	Valor x_i que más se repite	Valor X_i que más se repite
Medidas de localización o posición		
	$Población\ ordenada = (x_1, x_2, .., x_N)$	$Muestra\ ordenada = (X_1, X_2, .., X_n)$
Cuantil p	Si pN no es un número entero: $q_p = x_{\lfloor pN \rfloor + 1}$ Si pN es un número entero: $q_p = \frac{x_{pN} + x_{pN+1}}{2}$	Si pn no es un número entero: $\hat{q}_p = x_{\lfloor pn \rfloor + 1}$ Si pn es un número entero: $\hat{q}_p = \frac{X_{pn} + X_{pn+1}}{2}$
Percentil γ	Si $\gamma N/100$ no es un número entero: $\pi_\gamma = x_{\lfloor \gamma N/100 \rfloor + 1}$ Si $\gamma N/100$ es un número entero: $\pi_\gamma = \frac{x_{\gamma N/100} + x_{\gamma N/100+1}}{2}$	Si $\gamma n/100$ no es un número entero: $\hat{\pi}_\gamma = X_{\lfloor \gamma n/100 \rfloor + 1}$ Si $\gamma n/100$ es un número entero: $\hat{\pi}_\gamma = \frac{X_{\gamma n/100} + X_{\gamma n/100+1}}{2}$
Primer cuartil	Si $N/4$ no es un número entero: $Q_1 = x_{\lfloor N/4 \rfloor + 1}$ Si $N/4$ es un número entero: $Q_1 = \frac{x_{N/4} + x_{N/4+1}}{2}$	Si $n/4$ no es un número entero: $\hat{Q}_1 = X_{\lfloor N/4 \rfloor + 1}$ Si $n/4$ es un número entero: $\hat{Q}_1 = \frac{X_{n/4} + X_{n/4+1}}{2}$
Segundo cuartil	$Q_2 = M$	$\hat{Q}_2 = \widehat{M}$
Tercer cuartil	Si $3N/4$ no es un número entero: $Q_3 = x_{\lfloor 3N/4 \rfloor + 1}$ Si $3N/4$ es un número entero: $Q_3 = \frac{x_{3N/4} + x_{3N/4+1}}{2}$	Si $3n/4$ no es un número entero: $\hat{Q}_3 = X_{\lfloor 3N/4 \rfloor + 1}$ Si $3n/4$ es un número entero: $\hat{Q}_3 = \frac{X_{3n/4} + X_{3n/4+1}}{2}$

$\lfloor número \rfloor$ = parte entera de $número$

Medidas de dispersión o variabilidad		
Valor mínimo, valor máximo, rango	Si la población está ordenada de menor a mayor: $\min x = x_1$, $\max x = x_N$ $rango\ x = x_N - x_1$	Si la muestra está ordenada de menor a mayor: $\min X = X_1$, $\max X = X_n$ $rango\ X = X_n - X_1$
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	$IQR = Q_3 - Q_1$ Los datos atípicos están fuera del intervalo: $[Q_1 - 1.5 \cdot IQR,\ Q_3 + 1.5 \cdot IQR]$	$\widehat{IQR} = \hat{Q}_3 - \hat{Q}_1$ Los datos atípicos están fuera del intervalo: $\left[\hat{Q}_1 - 1.5 \cdot \widehat{IQR},\ \hat{Q}_3 + 1.5 \cdot \widehat{IQR}\right]$
Medidas de forma		
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		
Proporción de elementos que cumplen una condición	$Población = (x_1, x_2, .., x_N)$ $p = \frac{nº\ elementos\ que\ cumplen\ la\ condición}{N}$	$Muestra\ = (X_1, X_2, .., X_n)$ $\hat{p} = \frac{nº\ elementos\ que\ cumplen\ la\ condición}{n}$

Medidas estadísticas a partir de una tabla de frecuencias

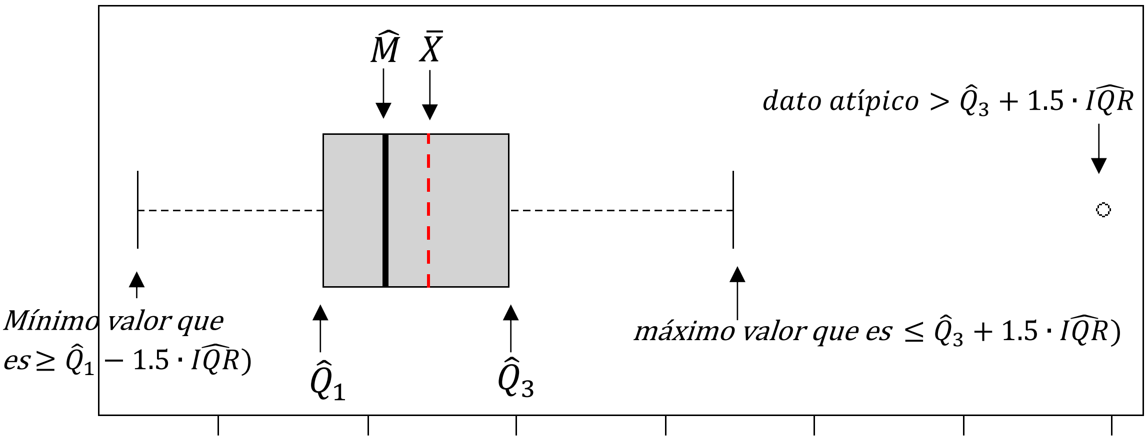
Fila	**x o X**	**f**
1	v_1	f_1
2	v_2	f_2
…	…	…
i	v_i	f_i
…	…	…
k	v_k	f_k

Fila	**x ó X**	**Marca**	**f**
1	$[a_1, b_1)$	m_1	f_1
2	$[a_2, b_2)$	m_2	f_2
…	…	…	…
i	$[a_i, b_i)$	m_i	f_i
…	…	…	…
k	$[a_k, b_k]$	m_k	f_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)

Medida	***Poblacional***	***Muestral***
Media (aritmética)	$\mu = \frac{\sum_{i=1}^{k} f_i v_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{k} f_i v_i}{n}$
	$\mu = \frac{\sum_{i=1}^{k} f_i m_i}{N}$ $m_i = \frac{a_i + b_i}{2}$	$\bar{X} = \frac{\sum_{i=1}^{k} f_i m_i}{n}$
Varianza	$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (v_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{k} f_i (v_i - \bar{X})^2}{n-1}$
	$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \bar{X})^2}{n-1}$

Diagrama de caja para una muestra



Estadística bidimensional

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

	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	h_w	...	h_{w_j}	...	h_{w_l}		1

Medida	*Poblacional*	*Muestral*
Medias (aritméticas) marginales	$\mu_x = \frac{\sum_{i=1}^{k} f_{v_i} v_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^{k} f_{v_i} v_i}{n}$
	$\mu_y = \frac{\sum_{i=1}^{l} f_{w_i} w_i}{N}$	$\bar{Y} = \frac{\sum_{i=1}^{l} f_{w_i} w_i}{n}$
Varianzas marginales	${\sigma_x}^2 = \frac{\sum_{i=1}^{k} f_{v_i}(v_i - \mu_x)^2}{N}$	${s_X}^2 = \frac{\sum_{i=1}^{k} f_{v_i}(v_i - \bar{X})^2}{n-1}$
	${\sigma_y}^2 = \frac{\sum_{i=1}^{k} f_{w_i}(w_i - \mu_y)^2}{N}$	${s_Y}^2 = \frac{\sum_{i=1}^{k} f_{w_i}(w_i - \bar{Y})^2}{n-1}$
Media de x (ó X) cuando y (ó Y) = w_j	$\mu_x \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} v_i}{f_{w_j}}$	$\bar{X} \vert w_j = \frac{\sum_{i=1}^{k} f_{i,j} v_i}{f_{w_j}}$
Media de y (ó Y) cuando x (ó X) = v_i	$\mu_y \vert v_i = \frac{\sum_{j=1}^{l} f_{i,j} w_j}{f_{v_i}}$	$\bar{Y} \vert v_i = \frac{\sum_{j=1}^{l} f_{i,j} w_j}{f_{v_i}}$