Formulario Estadística I

DISTRIBUCIÓN DE FRECUENCIAS DE DATOS NO AGRUPADOS EN INTERVALOS

Datos (x_i)	n_i	N_i	f_i	F_i
x_1	n_1	$N_1 = n_1$	$f_1 = n_1/N$	$F_1 = f_1 = N_1/N$
x_2	n_2	$N_2 = n_1 + n_2 = N_1 + n_2$	$f_2 = n_2/N$	$F_2 = f_1 + f_2 = F_1 + f_2 = N_2/N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_i	$N_i = n_1 + \cdots + n_i = N_{i-1} + n_i$	$f_i = n_i/N$	$F_i = f_1 + \cdots + f_i = F_{i-1} + f_i = N_i/N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	$N_k = n_1 + \cdots + n_k = N_{k-1} + n_k = N$	$f_k = n_k/N$	$F_k = f_1 + \cdots + f_k = F_{k-1} + f_k = N_k/N = 1$
	$\sum_{i=1}^{k} n_i = N$		$\sum_{i=1}^{k} f_i = 1$	

DISTRIBUCIÓN DE FRECUENCIAS DE DATOS AGRUPADOS EN INTERVALOS

Intervalos ($L_{i-1} - L_i$)	Marca de clase (x_i)	n_i	N_i	f_i	F_i	c_i	d_i
$L_0 - L_1$	$x_1 = (L_0 + L_1)/2$	n_1	N_1	f_1	F_1	$c_1 = L_1 - L_0$	$d_1 = n_1/c_1$
$L_1 - L_2$	$x_2 = (L_1 + L_2)/2$	n_2	N_2	f_2	F_2	$c_2 = L_2 - L_1$	$d_2 = n_2/c_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$L_{i-1} - L_i$	$x_i = (L_{i-1} + L_i)/2$	n_i	N_i	f_i	F_i	$c_i = L_i - L_{i-1}$	$d_i = n_i/c_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$L_{k-1} - L_k$	$x_k = (L_{k-1} + L_k)/2$	n_k	$N_k = N$	f_k	$F_k = 1$	$c_k = L_k - L_{k-1}$	$d_k = n_k/c_k$
		$\sum_{i=1}^{k} n_i = N$		$\sum_{i=1}^{k} f_i = 1$			

TABLA DE DOBLE ENTRADA O TABLA DE CORRELACIÓN

$x_i \backslash y_j$	y_1	$\dots$	y_j	$\dots$	y_h	$n_{i.}$
x_1	n_{11}	$\dots$	n_{1j}	$\dots$	n_{1h}	$n_{1.} = n_{11} + \dots + n_{1h}$
x_2	n_{21}	$\dots$	n_{2j}	$\dots$	n_{2h}	$n_{2.} = n_{21} + \dots + n_{2h}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	$\dots$	n_{ij}	$\dots$	n_{ih}	$n_{i.} = n_{i1} + \dots + n_{ih}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	n_{k1}	$\dots$	n_{kj}	$\dots$	n_{kh}	$n_{k.} = n_{k1} + \dots + n_{kh}$
$n_{.j}$	$n_{.1} = n_{11} + \dots + n_{k1}$	$\dots$	$n_{.j} = n_{1j} + \dots + n_{kj}$	$\dots$	$n_{.h} = n_{1h} + \dots + n_{kh}$	$\sum_{i=1}^{k} \sum_{j=1}^{h} n_{ij} = N$

TABLA DE CONTINGENCIA

$a_i \backslash b_j$	b_1	$\dots$	b_j	$\dots$	b_h	$n_{i.}$
a_1	n_{11}	$\dots$	n_{1j}	$\dots$	n_{1h}	$n_{1.} = n_{11} + \dots + n_{1h}$
a_2	n_{21}	$\dots$	n_{2j}	$\dots$	n_{2h}	$n_{2.} = n_{21} + \dots + n_{2h}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
a_i	n_{i1}	$\dots$	n_{ij}	$\dots$	n_{ih}	$n_{i.} = n_{i1} + \dots + n_{ih}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
a_k	n_{k1}	$\dots$	n_{kj}	$\dots$	n_{kh}	$n_{k.} = n_{k1} + \dots + n_{kh}$
$n_{.j}$	$n_{.1} = n_{11} + \dots + n_{k1}$	$\dots$	$n_{.j} = n_{1j} + \dots + n_{kj}$	$\dots$	$n_{.h} = n_{1h} + \dots + n_{kh}$	$\sum_{i=1}^{k} \sum_{j=1}^{h} n_{ij} = N$

MEDIDAS DE POSICIÓN, DISPERSIÓN, FORMA Y CONCENTRACIÓN

Media aritmética	$\bar{x} = \frac{1}{N}\sum_{i=1}^{k} x_i n_i$		
Media aritmética ponderada	$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \ldots + w_kx_k}{w}$		
Media de la composición de poblaciones	$\bar{x}_p = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + \ldots + N_k\bar{x}_k}{N}$		
Media geométrica	$\bar{x}_g = \sqrt[N]{x_1^{n_1}\cdots x_k^{n_k}} = (x_1^{n_1}\cdots x_k^{n_k})^{1/N}$		
Media armónica	$\bar{x}_a = \frac{N}{\frac{n_1}{x_1}+\frac{n_2}{x_2}+\ldots+\frac{n_k}{x_k}}$		
Mediana	Datos no agrupados en intervalos	$N_i = N/2$	$Me = (x_i + x_{i+1})/2$
		$N_i > N/2$	$Me = x_i$
	Datos agrupados en intervalos	$N_i = N/2$	$Me = L_i$
		$N_i > N/2$	$Me = L_{i-1} + \frac{N/2 - N_{i-1}}{n_i}\cdot c_i$
Moda	Datos agrupados en intervalos	$Mo = L_{i-1} + \frac{n_{i+1}}{n_{i-1}+n_{i+1}}\cdot c_i$	
		$Mo = L_{i-1} + \frac{d_{i+1}}{d_{i-1}+d_{i+1}}\cdot c_i$	
Cuantiles Cuartiles (Q_r): $k = 4$, $r = 1, 2, 3$ Deciles (D_r): $k = 10$, $r = 1,\ldots, 9$ Percentiles (P_r): $k = 100$, $r = 1,\ldots,99$ $r = \frac{100}{N}[(P_r - L_{i-1})\frac{n_i}{c_i} + N_{i-1}]$	Datos no agrupados en intervalos	$N_i = (r/k)\cdot N$	$C_{r/k} = \frac{x_i + x_{i+1}}{2}$
		$N_i > (r/k)\cdot N$	$C_{r/k} = x_i$
	Datos agrupados en intervalos	$N_i = (r/k)\cdot N$	$C_{r/k} = L_i$
		$N_i > (r/k)\cdot N$	$C_{r/k} = L_{i-1} + \frac{(r/k)\cdot N - N_{i-1}}{n_i}\cdot c_i$

Rango	$R = \text{máx}\{x_1,\ldots,x_k\} - \text{mín}\{x_1,\ldots,x_k\}$
Recorrido intercuartílico	$R_I = Q_3 - Q_1$
Desviación absoluta media	$D_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{k} \lvert x_i - \bar{x} \rvert n_i$
Varianza	$S^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \frac{1}{N}\sum_{i=1}^{k} x_i^2 n_i - \bar{x}^2$
Desviación típica	$S = +\sqrt{S^2}$
Coeficiente de apertura	$A = \text{máx}\{x_1,\ldots,x_k\}/\text{mín}\{x_1,\ldots,x_k\}$
Recorrido relativo	$R_R = [\text{máx}\{x_1,\ldots,x_k\} - \text{mín}\{x_1,\ldots,x_k\}]/\bar{x}$
Recorrido semi-interc.	$R_S = (Q_3 - Q_1)/(Q_1 + Q_3)$
Coeficiente de variación	$CV = S / \lvert \bar{x} \rvert$
Variable tipificada	$Z = (X - \bar{x})/S$
Coeficiente de asimetría de Fisher	$g_1 = \frac{m_3}{S^3} = \frac{\frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i}{(S^2)^{3/2}}$
Coeficiente de asimetría de Yule-Bowley	$AB = (Q_1 + Q_3 - 2Q_2)/(Q_3 - Q_1)$
Coeficiente de curtosis	$g_2 = \frac{m_4}{S^4} - 3 = \frac{\frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i}{(S^2)^2} - 3$
Índice de Gini	$IG = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$ $p_i = \frac{N_i}{N}\cdot 100$; $q_i = \frac{u_i}{u_k}\cdot 100$

DISTRIBUCIONES BIDIMENSIONALES Y REGRESIÓN

Independencia estadística	$n_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{N} \quad \forall\, i,j$	**Covarianza**	$S_{XY} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{h}(x_i - \bar{x})(y_j - \bar{y})n_{ij} = \frac{1}{N}\sum_{i=1}^{k} x_i \sum_{j=1}^{h} y_j n_{ij} - \bar{x}\bar{y}$	
Coeficiente de correlación lineal	$r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y}$	**Rectas de regresión lineal**	$y - \bar{y} = \frac{S_{XY}}{S_X^2}(x - \bar{x})$	$x - \bar{x} = \frac{S_{XY}}{S_Y^2}(y - \bar{y})$
Coeficiente de determinación	$R_{XY}^2 = r_{XY}^2$			

ATRIBUTOS

Coeficiente básico de dependencia	$D = n_{11} - \frac{n_{1\cdot} \times n_{\cdot 1}}{N} = \frac{n_{11}n_{22} - n_{12}n_{21}}{N}$	**Coeficiente de asociación *Q* de Yule**	$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}} = \frac{N \cdot D}{n_{11}n_{22} + n_{12}n_{21}}$

Estadístico chi-cuadrado	$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{h}\frac{(n_{ij} - e_{ij})^2}{e_{ij}}$	**Coeficiente de contingencia**	$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$	**Coeficiente *V* de Cramer**	$V = \sqrt{\frac{\chi^2}{N[m\acute{\imath}n(k,h) - 1]}}$	**Coeficiente *T* de Tschuprow**	$T = \frac{\chi^2}{N\sqrt{(k-1)(h-1)}}$

Coeficiente de correlación por rangos de Spearman	$\rho = 1 - \frac{6}{N(N^2 - 1)}\sum_{i=1}^{n} d_i^2$

ÍNDICES Y TASAS DE VARIACIÓN

Índices simples	$I_{t/0} = (x_t / x_0) \cdot 100$	**Tasa de variación absoluta**	$\nabla x_t = x_t - x_{t-1}$
Índices en cadena	$I_{t/t-1} = (x_t / x_{t-1}) \cdot 100$	**Tasa de variación relativa**	$\dot{x}_t = \left(\frac{x_t - x_{t-1}}{x_{t-1}}\right) \cdot 100 = \left(\frac{x_t}{x_{t-1}} - 1\right) \cdot 100$
Cambio de base	$I_{t/t''} = \frac{I_{t/t'}}{I_{t''/t'}}$	**Tasa media acumulativa**	$r_{t_1/t_2} = \left[(x_{t_2} / x_{t_1})^{\frac{1}{t_2 - t_1}} - 1\right] \cdot 100$
Deflactación	Magnitud del año t a precios constantes (base 0) $= \frac{\text{Magnitud del año } t \text{ a precios corrientes}}{I_{t/0}}$		

PROBABILIDAD

Probabilidades hipergeométricas	$\Pr(k \text{ bolas blancas}) = \frac{\binom{N_1}{k} \cdot \binom{N_2}{n-k}}{\binom{N}{n}}$	**Teorema de Bayes**	$\Pr(B_k \mid A) = \frac{\Pr(A \mid B_k) \cdot \Pr(B_k)}{\sum_{k=1}^{n}\Pr(A \mid B_k) \cdot \Pr(B_k)}$
Probabilidad condicionada	$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$	**Teorema de la probabilidad total**	$\Pr(A) = \sum_{k=1}^{n}\Pr(A \mid B_k) \cdot \Pr(B_k)$