1. DISTRIBUCIONES BIDIMENSIONALES

Los N elementos de la población son clasificados según dos caracteres X e Y, cuyas modalidades las notamos respectivamente por x_i e y_j donde i varía desde 1 hasta p y j varía desde 1 hasta q.

n_{ij} → **Frecuencia absoluta**: número de individuos de la población que presentan la modalidad x_i de X y la modalidad y_j de Y .

Tabla de frecuencias absolutas de una variable bidimensional

	(X/Y)	y_1	y_2	y_3	…	y_q	suma $n_{i\cdot}$
	x_1	n_{11}	n_{12}	n_{13}	…	n_{1q}	$n_{1\cdot}$
	x_2	n_{21}	n_{22}	n_{23}	…	n_{2q}	$n_{2\cdot}$
	x_3	n_{31}	n_{32}	n_{33}	…	n_{3q}	$n_{3\cdot}$
	…	…	…	…	…	…	…
	x_p	n_{p1}	n_{p2}	n_{p3}	…	n_{pq}	$n_{p\cdot}$
suma	$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	…	$n_{\cdot q}$	$n_{\cdot\cdot}=N$

Nota: el primer subíndice indica el número de fila y el segundo subíndice indica el número de columna

f_{ij} → **Frecuencia relativa**: proporción de individuos de la población que presentan la modalidad x_i de X y la modalidad y_j de Y. Se obtiene dividiendo la frecuencia absoluta entre el número de elementos de la población (N).

Tabla de frecuencias relativas de una variable bidimensional

	(X/Y)	y_1	y_2	y_3	…	y_q	suma $n_{i\cdot}$
	x_1	f_{11}	f_{12}	f_{13}	…	f_{1q}	$f_{1\cdot}$
	x_2	f_{21}	f_{22}	f_{23}	…	f_{2q}	$f_{2\cdot}$
	x_3	f_{31}	f_{32}	f_{33}	…	f_{3q}	$f_{3\cdot}$
	…	…	…	…	…	…	…
	x_p	f_{p1}	f_{p2}	f_{p3}	…	f_{pq}	$f_{p\cdot}$
suma	$f_{\cdot j}$	$f_{\cdot 1}$	$f_{\cdot 2}$	$f_{\cdot 3}$	…	$f_{\cdot q}$	$f_{\cdot\cdot}=1$

2. DISTRIBUCIONES MARGINALES

Estas distribuciones nos indican cómo se distribuye una variable independientemente de los valores que tome la otra.

Distribución marginal de X:

X	$n_{i\cdot}$
x_1	$n_{1\cdot}$
x_2	$n_{2\cdot}$
x_3	$n_{3\cdot}$
…	…
x_p	$n_{p\cdot}$
suma	$n_{\cdot\cdot}=N$

Distribución marginal de Y:

Y	$n_{\cdot j}$
y_1	$n_{\cdot 1}$
y_2	$n_{\cdot 2}$
y_3	$n_{\cdot 3}$
…	…
y_q	$n_{\cdot q}$
suma	$n_{\cdot\cdot}=N$

Existe una única distribución marginal de X y una única distribución marginal de Y.

3. DISTRIBUCIONES CONDICIONADAS

Son distribuciones unidimensionales obtenidas a partir de las bidimensionales, manteniendo fijo un valor en una de las variables y considerando los valores que toma la otra con sus respectivas frecuencias

Distribución de *X* condicionada al valor y_j de *Y*:

$X\|Y=y_j$	$n_{i\|j}$
x_1	n_{1j}
x_2	n_{2j}
x_3	n_{3j}
...	...
x_p	n_{pj}
suma	$n_{\cdot j}$

Distribución de *Y* condicionada al valor x_i de *X*:

$Y\|X=x_i$	$n_{j\|i}$
y_1	n_{i1}
y_2	n_{i2}
y_3	n_{i3}
...	...
y_q	n_{iq}
suma	$n_{i\cdot}$

4. COVARIANZA

$$\sigma_{XY} = Cov(X,Y) = \frac{1}{N}\sum_{j=1}^{q}\sum_{i=1}^{p}(x_i-\bar{x})(y_j-\bar{y})n_{ij} = \frac{1}{N}\sum_{j=1}^{q}\sum_{i=1}^{p}x_i y_j n_{ij} - \overline{x}\overline{y}$$

5. INDEPENDENCIA ESTADÍSTICA

- Las variables *X* e *Y* se dicen **estadísticamente dependientes** cuando la variación de una influye en la distribución de frecuencias de la otra.

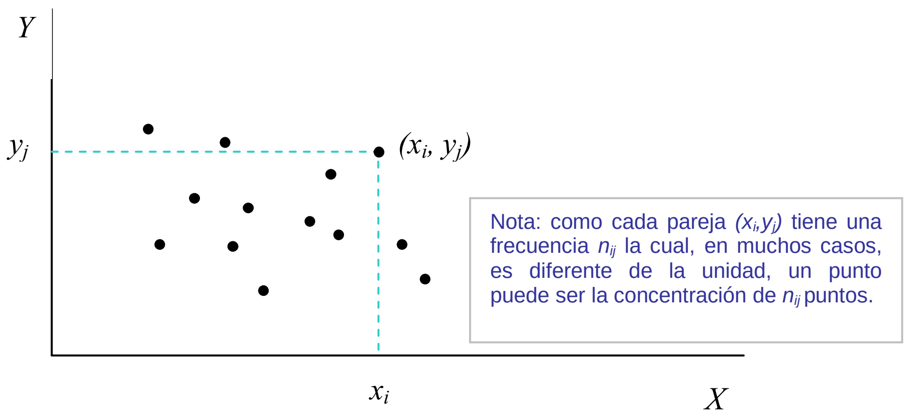
- Cuando las distribuciones condicionadas a cualquier valor de la otra variable sean iguales, diremos que las variables X e Y son estadísticamente **independientes**. Las variables son independientes si y solamente si ocurre que $f_{ij} = f_{i\cdot} f_{\cdot j}$ para cualquier pareja *(i,j)*.

- Cuando las variables son independientes la covarianza vale 0 (siempre).

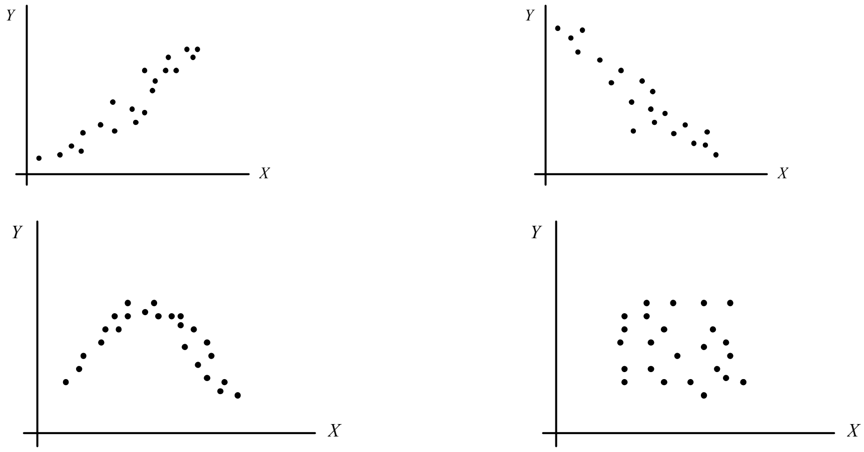
- Cuando la covarianza vale 0 las variables son **incorreladas**.

6. NUBE DE PUNTOS:

Se representan las observaciones en unos ejes cartesianos. Para cada individuo se tiene un punto con coordenadas (x_i,y_j), que respresenta el valor observado en *(X, Y)*.



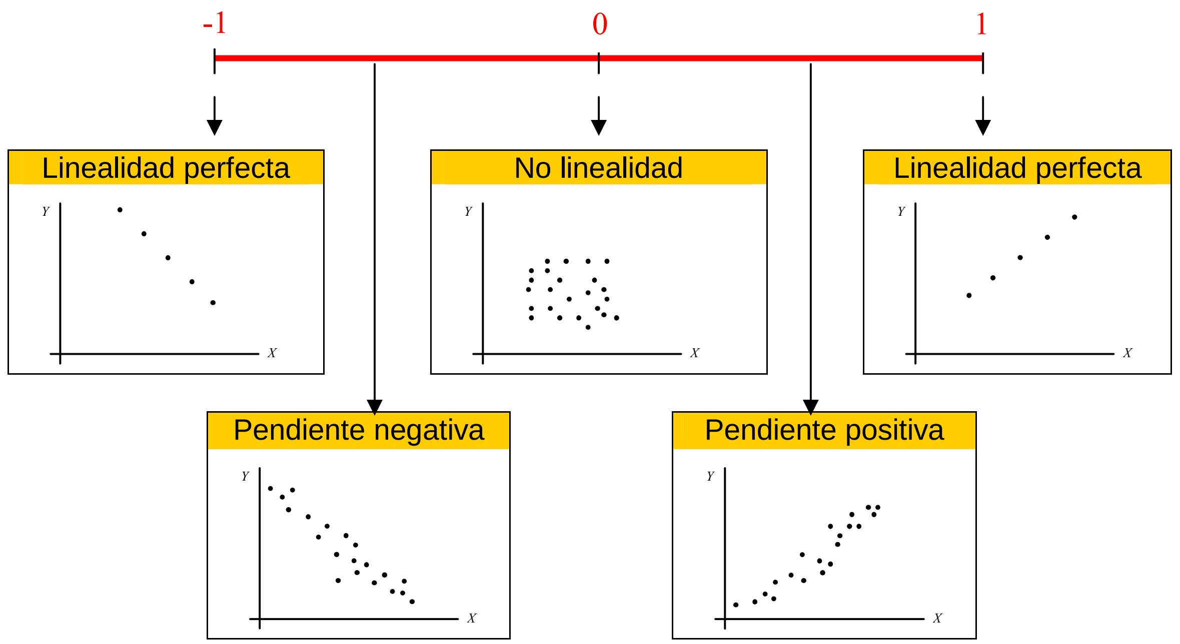
Algunos ejemplos:

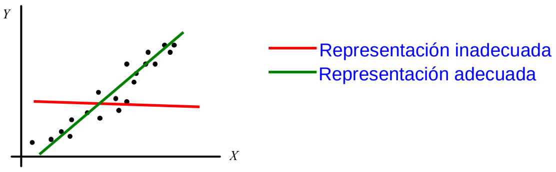


7. RECTAS DE REGRESIÓN:

COEFICIENTE de CORRELACIÓN LINEAL de PEARSON: $r = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Valores:

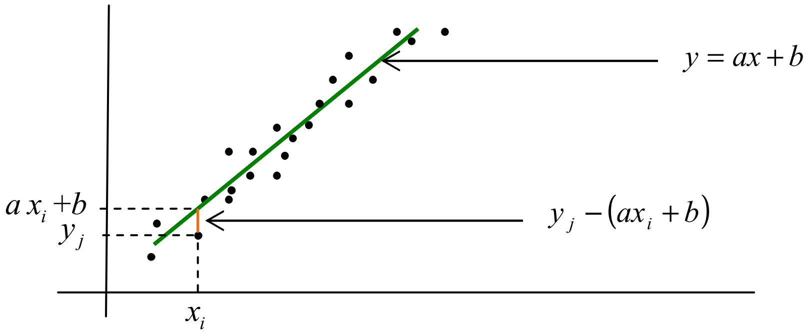




Para que una recta represente bien a los puntos, debe existir linealidad y debemos utilizar un método que reduzca la distancia entre los puntos observados y la recta.

Método de los **mínimos cuadrados**: se minimiza la suma de los residuos al cuadrado.
Para predecir Y en función de X: $Y = aX + b$

$$\min\left(\frac{1}{n}\sum_{i,j}\left(y_j - (ax_i + b)\right)^2 n_{ij}\right)$$



Y|*X*	*X*|*Y*
Variable dependiente: *Y* Variable independiente: *X*	Variable dependiente: *X* Variable independiente: *Y*
Recta de regresión: $Y = aX + b$	Recta de regresión: $X = \alpha Y + \beta$
$a = \frac{\sigma_{XY}}{\sigma_X^2}$ $\quad b = \overline{y} - \frac{\sigma_{XY}}{\sigma_X^2}\overline{x}$	$\alpha = \frac{\sigma_{XY}}{\sigma_Y^2}$ $\quad \beta = \overline{x} - \frac{\sigma_{XY}}{\sigma_Y^2}\overline{y}$

Posición relativa de las rectas de regresión:

Ambas rectas pasan por el punto $(\overline{x}, \overline{y})$
- Si las pendientes de las rectas no son iguales, las rectas se cortan en ese punto.
- Si las pendientes de las rectas son iguales, las rectas son coincidentes.

8. BONDAD DE AJUSTE

COEFICIENTE de DETERMINACIÓN: $R^2 = 1 - \frac{S_{rY}^2}{\sigma_Y^2}$

Valores: