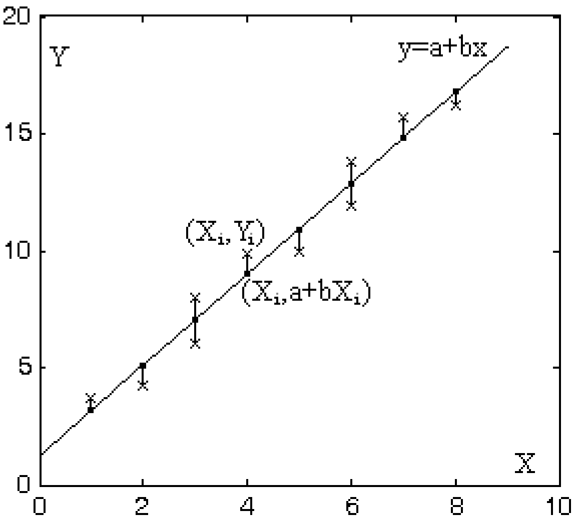
1.4. Obtener una recta que tenga la menor desviación posible de un conjunto de puntos:

Objetivos:

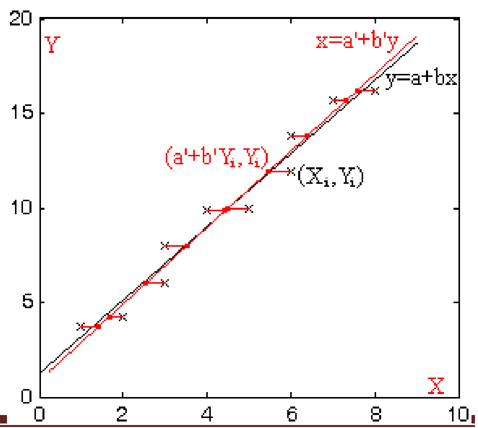
1. Obtener la recta que minimice la suma de las desviaciones cuadráticas de las ordenadas de un conjunto de puntos.
2. Obtener la recta que minimice la suma de las desviaciones cuadráticas de las abscisas de un conjunto de puntos.
3. Valorar el grado de ajuste de la recta de regresión al correspondiente conjunto de puntos.

Actividad 1.36. Si tenemos un conjunto de puntos $\{X_i, Y_i\}_{i=1\ldots n}$, diremos que y=a+bx es la recta *de regresión* de Y sobre X si y sólo si
$\sum_{i=1}^{n}(Y_i-(a+bX_i))^2$ es mínimo.
Teniendo la cuenta el
Teorema -1.8: si una función derivable f(x,y) tiene un mínimo en (a,b), entonces $f_x'(a,b)=0$ y $f_y'(a,b)=0$.
demostrar
Teorema 1.31: si y=a+bx es la recta de regresión de Y sobre X, entonces $a+b\cdot\mu(X)=\mu(Y)$, $a\cdot\mu(X)+b\cdot\mu(X^2)=\mu(XY)$.
Teorema 1.32: si y=a+bx es la recta de regresión de Y sobre X, entonces $b=c_{XY}/\sigma(X)^2$, $a=\mu(Y)-b\cdot\mu(X)$, dónde $c_{XY}=\mu(XY)-\mu(X)\cdot\mu(Y)$ (*covarianza* de X y Y).



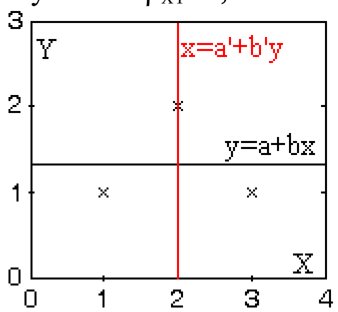
Actividad 1.37. Teniendo en cuenta el
Teorema -1.9: si para una función f(x,y) derivable hasta segundo orden se cumple $f_x'(a,b)=0$, $f_y'(a,b)=0$, $f_{xx}''(a,b)>0$, $f_{xy}''(a,b)^2<f_{xx}''(a,b)\cdot f_{yy}''(a,b)$, entonces f(x,y) tiene un mínimo en (a,b)
demostrar el
Teorema 1.33: si $\sigma(X)^2>0$, $b=c_{XY}/\sigma(X)^2$, entonces $y-\mu(Y)=b\cdot(x-\mu(X))$ es la recta de regresión de Y sobre X.
Observemos que el "centro de masas" $(\mu(X),\mu(Y))$ pertenece siempre a la recta de regresión.
Problema 1.19: obtener la recta de regresión del número de calzado sobre la edad en el alumnado asistente a clase; valorarla.

Actividad 1.38. Intercambiando la X y la Y obtenemos el
Teorema 1.34: si $\sigma(Y)^2>0$, $b'=c_{XY}/\sigma(Y)^2$, entonces $x-\mu(X)=b\cdot(y-\mu(Y))$ es la recta de regresión de X sobre Y.



Si $\sigma(X)^2>0$ y $\sigma(Y)^2>0$, ambas rectas de regresión pasarán por el "centro de masas" $(\mu(X),\mu(Y))$, y definimos el *coeficiente de correlación* de X y Y por
$\rho_{XY} = c_{XY}/(\sigma(X)\sigma(Y))$.
Demostrar
Teorema 1.35: las rectas de regresión de Y sobre X y de X sobre Y coinciden si y sólo si $\rho_{XY} = \pm 1$.
Teorema 1.36: si $\rho_{XY} = 0$, entonces las rectas de regresión son $y=\mu(Y)$, $x=\mu(X)$ (perpendiculares).

Actividad 1.39. Diremos que dos variables aleatorias X, Y no tienen correlación lineal si y sólo si $\rho_{XY}=0$; esta condición es equivalente a la de $c_{XY}=0$ con $\sigma(X)>0$ y $\sigma(Y)>0$.



Demostrar el
Teorema 1.36: si dos variables aleatorias son independientes, no tienen correlación lineal.
¿La recíproca es cierta? Comprobarlo en el siguiente
Problema 1.20: estudiar la correlación lineal en el caso

X	1	2	3
Y	1	2	1

¿X e Y son independientes?

Actividad 1.40. Demostrar
Teorema 1.37: $\sigma(X\pm Y)^2 = \sigma(X)^2 + \sigma(Y)^2 \pm 2\cdot c_{XY}$.
Teorema 1.38: si X,Y son independientes o simplemente no tienen correlación lineal, entonces $\sigma(X\pm Y)^2 = \sigma(X)^2 + \sigma(Y)^2$.

Actividad 1.41. Demostrar, utilizando el Teorema 1.37,
Teorema 1.39: Si $y=a+bx$ es la recta de regresión de Y sobre X, entonces $\sigma(Y-bX)^2 = \sigma(Y)^2(1-\rho_{XY}{}^2)$.
Teorema 1.40: $-1 \le \rho_{XY} \le 1$.
Si $\rho_{XY}>0$ diremos que X,Y tienen correlación lineal positiva; si $\rho_{XY}<0$, diremos que X,Y tienen correlación lineal negativa; si $|\rho_{XY}|\approx 1$, diremos que X,Y tienen buena correlación lineal; si $\rho_{XY}\approx 0$, diremos que X,Y tienen mala correlación lineal.
Problema 1.21: estudiar la correlación lineal entre el número de calzado y la edad del alumnado asistente a clase; valorarla.