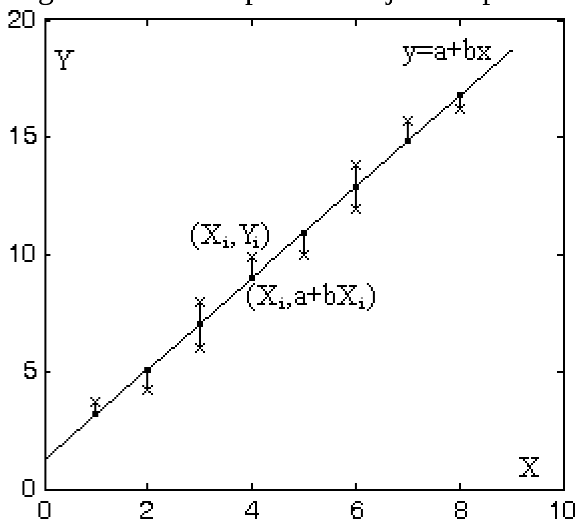
1.4. Obtenir una recta que tinga la menor desviació possible d'un conjunt de punts:

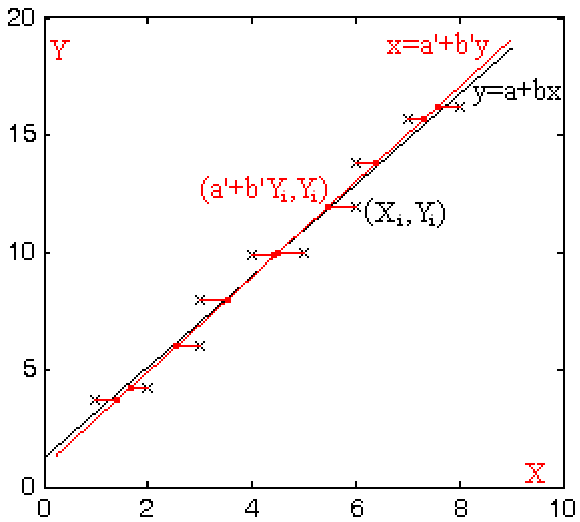
Objectius:

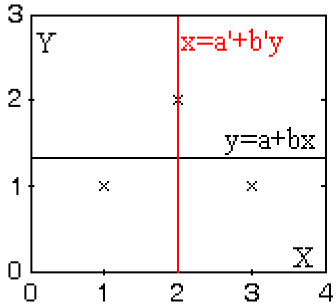
1. Obtenir la recta que minimitze la suma de les desviacions quadràtiques de les ordenades d'un conjunt de punts.
2. Obtenir la recta que minimitze la suma de les desviacions quadràtiques de les abcises d'un conjunt de punts.
3. Valorar el grau d'ajust de la recta de regressió al corresponent conjunt de punts.
4. **Activitat 1.36.** Si tenim un conjunt de punts $\{X_i, Y_i\}$i=1...n, direm que y=a+bx és la *recta de regressió* de Y sobre X si i sols si $\sum_{i=1}^{n}(Y_i-(a+bX_i))^2$ és mínim.
   

   Tenint el compte el
   Teorema -1.8: si una funció derivable f(x,y) té un mínim en (a,b), aleshores $f_x'(a,b)=0$ i $f_y'(a,b)=0$
   demostrar
   Teorema 1.31: si y=a+bx és la recta de regressió de Y sobre X, aleshores $a+b\cdot\mu(X)=\mu(Y)$, $a\cdot\mu(X)+b\cdot\mu(X^2)=\mu(XY)$.
   Teorema 1.32: si y=a+bx és la recta de regressió de Y sobre X, aleshores $b=c_{XY}/\sigma(X)^2$, $a=\mu(Y)-b\cdot\mu(X)$, on $c_{XY}=\mu(XY)-\mu(X)\cdot\mu(Y)$ (*covariança* de X i Y).

5. **Activitat 1.37.** Tenint en compte el
   Teorema -1.9: si per a una funció f(x,y) derivable fins a segon ordre s'acompleix $f_x'(a,b)=0$, $f_y'(a,b)=0$, $f_{xx}''(a,b)>0$, $f_{xy}''(a,b)^2<f_{xx}''(a,b)\cdot f_{yy}''(a,b)$, aleshores f(x,y) té un mínim en (a,b)
   demostrar el
   Teorema 1.33: si $\sigma(X)^2>0$, $b=c_{XY}/\sigma(X)^2$, aleshores $y-\mu(Y)=b\cdot(x-\mu(X))$ és la recta de regressió de Y sobre X.
   Observem que el "centre de mases" $(\mu(X),\mu(Y))$ pertany sempre a la recta de regressió.
   Problema 1.19: obtenir la recta de regresió del número de calcer sobre l'edat en l'alumnat assistent a classe; valorar-la.

6. **Activitat 1.38.** Intercanviant la X i la Y obtenim el
<u>Teorema 1.34:</u> si $\sigma(Y)^2>0$, $b'=c_{XY}/\sigma(Y)^2$, aleshores $x-\mu(X)=b\cdot(y-\mu(Y))$ és la recta de regressió de X sobre Y.
Si $\sigma(X)^2>0$ i $\sigma(Y)^2>0$, ambdues rectes de regressió passaran pel "centre de mases" $(\mu(X),\mu(Y))$, i definim el *coeficient de correlació* de X i Y per
$\rho_{XY} = c_{XY}/(\sigma(X)\sigma(Y))$.
Demostrar
<u>Teorema 1.35:</u> les rectes de regressió de Y sobre X i de X sobre Y coincideixen si i sols si $\rho_{XY} = \pm 1$.
<u>Teorema 1.36:</u> si $\rho_{XY} = 0$, aleshores les rectes de regressió són $y=\mu(Y)$, $x=\mu(X)$ (perpendiculars).





Activitat 1.39. Direm que dues variables aleatòries X, Y no tenen correlació lineal si i sols si $\rho_{XY}=0$; aquesta condició és equivalent a la de $c_{XY}=0$ amb $\sigma(X)>0$ i $\sigma(Y)>0$.
Demostrar el
<u>Teorema 1.36:</u> si dues variables aleatòries són independents, no tenen correlació lineal.
La recíproca és certa? Comprobar-ho en el següent

<u>Problema 1.20:</u> estudiar la correlació lineal en el cas

X	1	2	3
Y	1	2	1

X i Y són independents?

Activitat 1.40. Demostrar
<u>Teorema 1.37:</u> $\sigma(X\pm Y)^2 = \sigma(X)^2 + \sigma(Y)^2 \pm 2\cdot c_{XY}$.
<u>Teorema 1.38:</u> si X,Y són independents o simplement no tenen correlació lineal, aleshores $\sigma(X\pm Y)^2 = \sigma(X)^2 + \sigma(Y)^2$.

Activitat 1.41. Demostrar, utilitzant el Teorema 1.37,
<u>Teorema 1.39:</u> Si $y=a+bx$ és la recta de regressió de Y sobre X, aleshores $\sigma(Y-bX)^2 = \sigma(Y)^2(1-\rho_{XY}^2)$.
<u>Teorema 1.40:</u> $-1 \leq \rho_{XY} \leq 1$.
Si $\rho_{XY}>0$ direm que X,Y tenen correlació lineal positiva; si $\rho_{XY}<0$, direm que X,Y tenen correlació lineal positiva; si $|\rho_{XY}|\approx 1$, direm que X,Y tenen bona correlació lineal; si $\rho_{XY}\approx 0$, direm que X,Y tenen mala correlació lineal.
<u>Problema 1.21:</u> estudiar la correlació lineal entre el número de calcer i l'edat de l'alumnat assistent a classe; valorar-la.