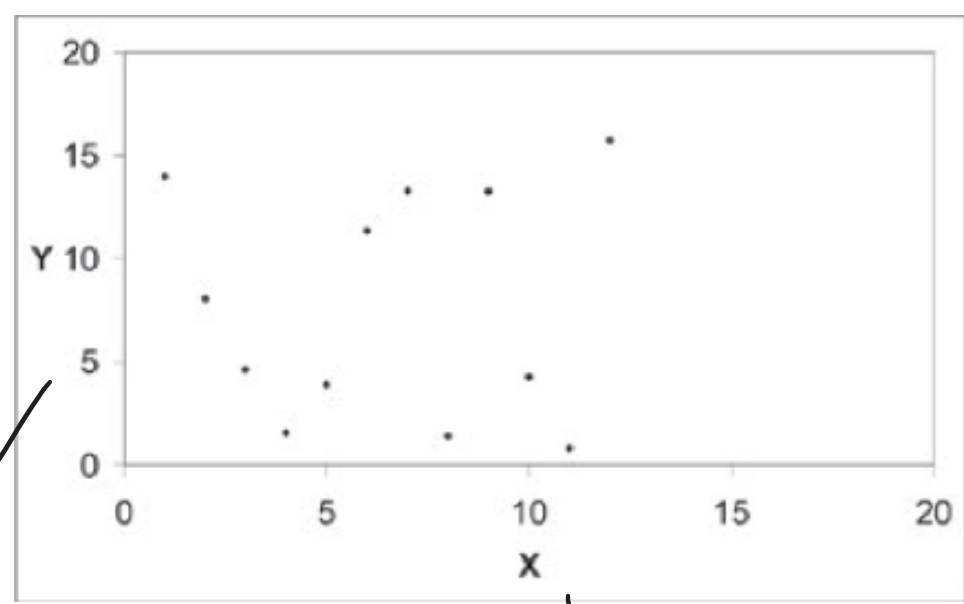
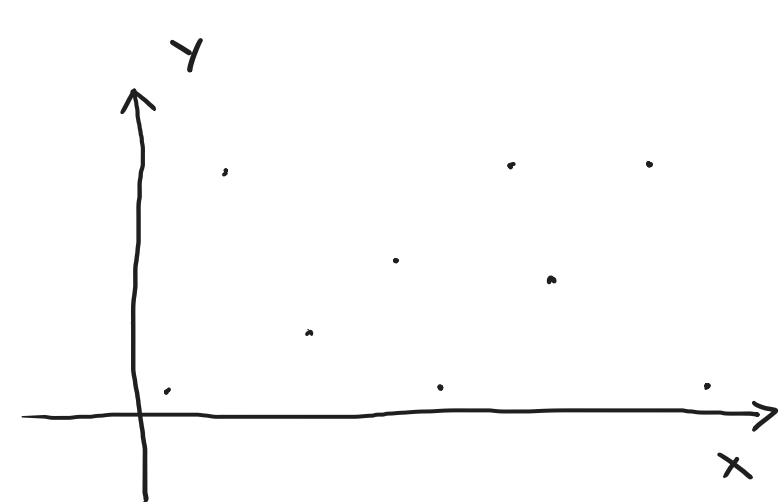
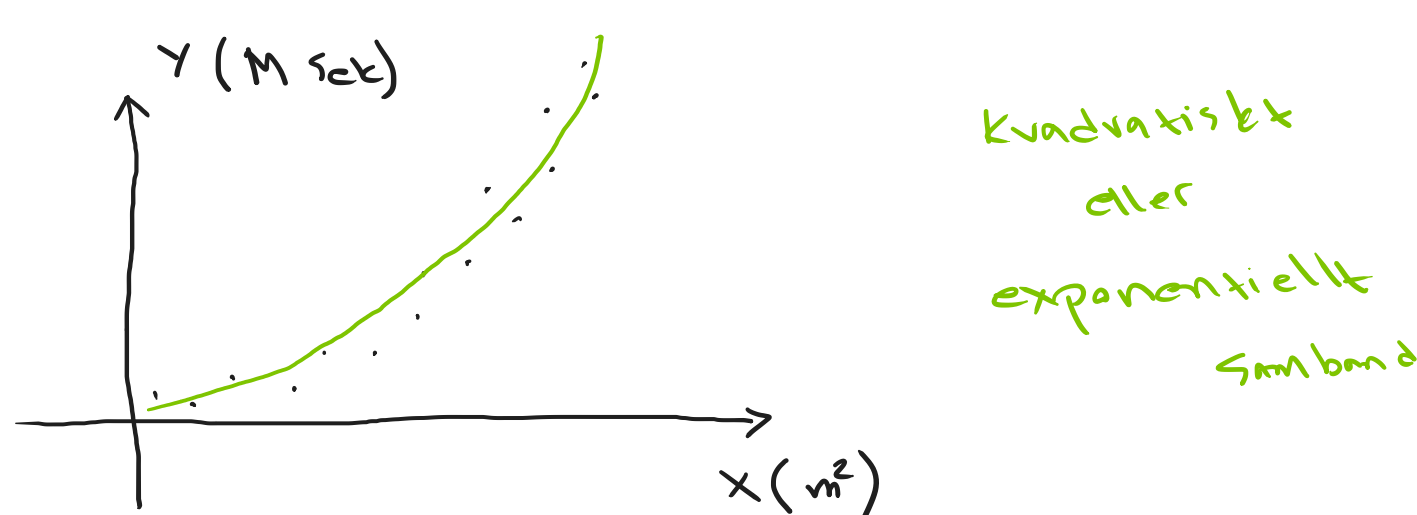
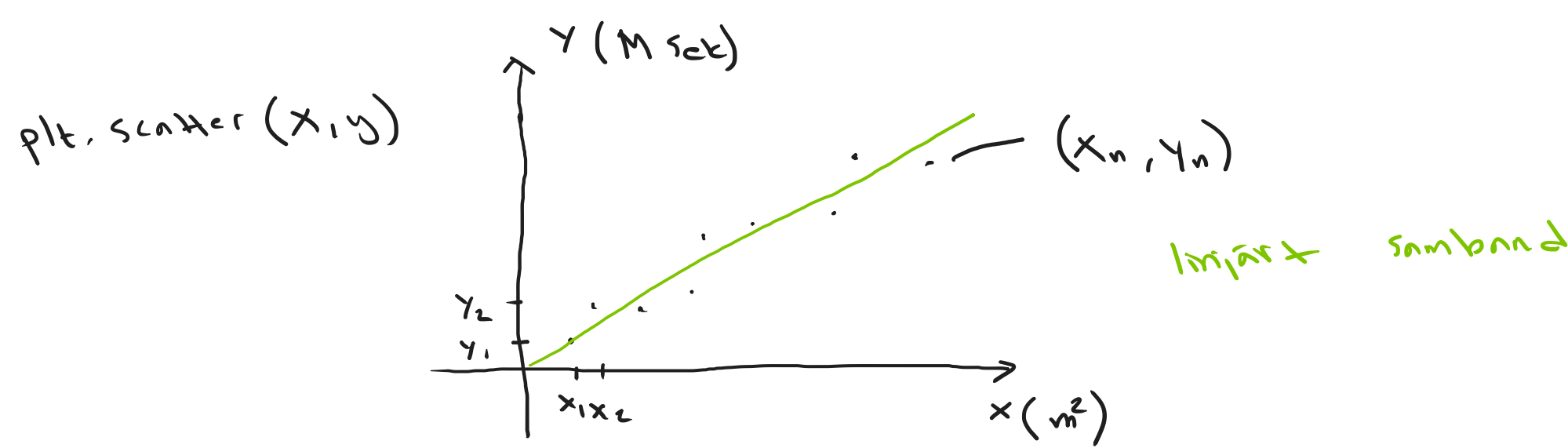


Korrelation

$x = [x_1, x_2, x_3, \dots, x_n]$   $n^2$  för varje bestånds  
 $y = [y_1, y_2, y_3, \dots, y_n]$  slutpris för samma bestånds



antal hg och storleken på var fast i Malmå ligger

Så, två variabler kan ha olika samband med varandra. Dessa samband kan faktiskt gestalta sig på oändligt många sätt.

Med hjälp av korrelationsanalyser kan vi mäta graden av samband mellan variabler.

Specifikt, kommer vi att här lära oss mäta hur starkt **linjärt samband** det är mellan två variabler

Korrelationskoefficienten

$$r = \frac{SS_{xy}}{\sqrt{SS_x \cdot SS_y}}$$

där

$$SS_x = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$SS_y = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SS_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

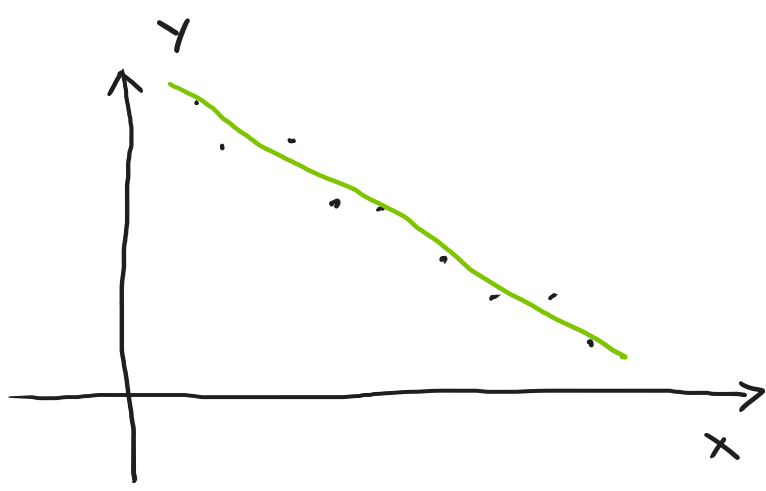
Det går nu att matematiskt visa följande

$$\rightarrow -1 \leq r \leq 1$$

$r = 1 \rightarrow$  perfekt positivt linjärt samband / korrelation

$r = 0 \rightarrow$  inget samband / korrelation

$r = -1 \rightarrow$  perfekt negativt linjärt samband / korrelation



$r > 0 \rightarrow$  positivt samband / korrelation

$r < 0 \rightarrow$  negativt samband / korrelation

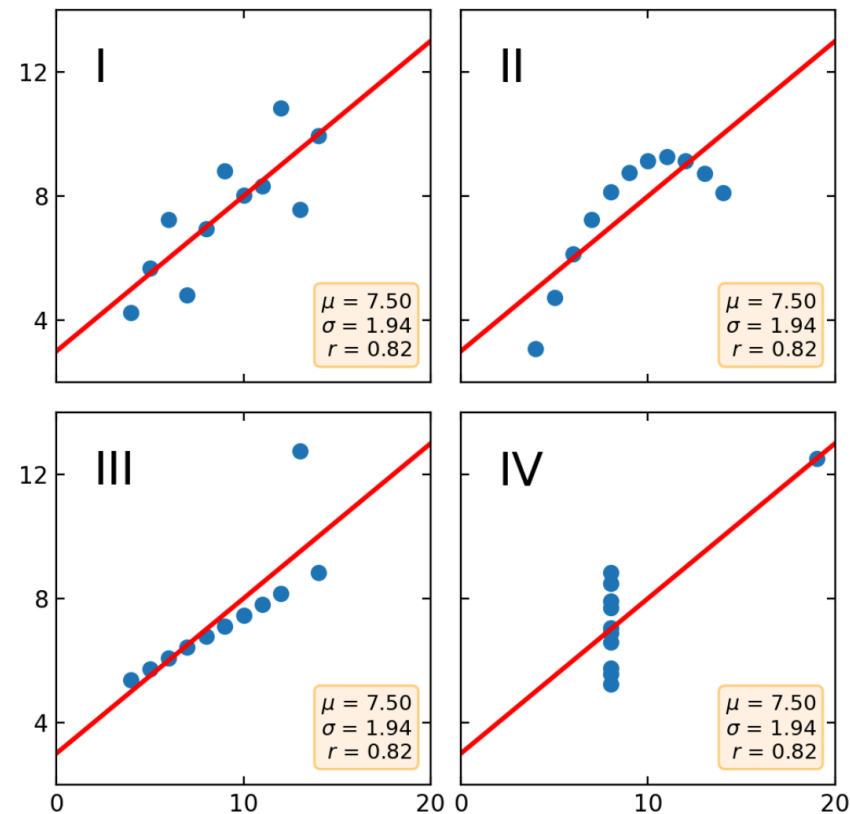
Uppgift

$$SS_x = \sum_{i=1}^n (x_i - \bar{x})^2$$
$$SS_y = \sum_{i=1}^n (y_i - \bar{y})^2$$
$$SS_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Definiera en funktion i Python som, givet två listor med siffror som argument, beräknar korrelationskoefficienten r.

Plotta därefter även x & y i en scatter plot - samt printa korrelationskoefficienten som titel

$$r = \frac{SS_{xy}}{\sqrt{SS_x \cdot SS_y}}$$



OBS!  
Var väldigt försiktig med hur ni tolkar korrelationskoefficienter!