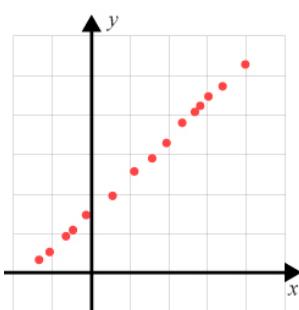


## Linjär regression

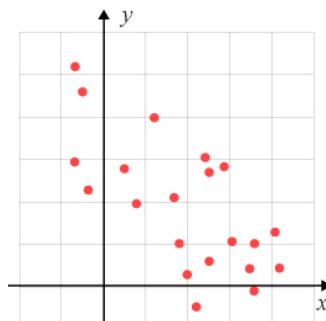
### Repetition från Matematik 1c

Om ett samband finns mellan två variabler har vi en **korrelation**. Huruvida ett samband finns, kan undersökas med ett **spridningsdiagram**. Om punkterna i ett spridningsdiagram ligger tydligt på en linje är korrelationen **stark**. Om punkterna ligger kring en rät linje, men spridda är korrelationen **svag**. Om punkterna ligger kring en linje med positiv lutning är korrelationen **positiv**. Om punkterna ligger kring en linje med negativ lutning är korrelationen **negativ**.

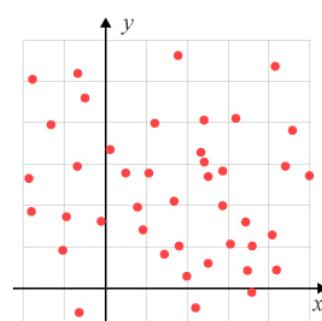
Linjen som värdena ligger samlade kring kallas **regressionslinje**. Om vi tar fram ekvationen för regressionslinjen gör vi en **linjär regression**. En regressionsanalys går att göra ungefärligt för hand eller exaktare med digitala verktyg. Vid regressionsanalysen kan vi ta fram ett värde som säger hur stark korrelationen är. Detta värde kallas **korrelationskoefficient**, betecknas  $r$  och är alltid ett värde mellan  $-1$  och  $1$ . Om punkterna ligger perfekt på en linje med positiv lutning är  $r = 1$ . Om punkterna ligger perfekt på en linje med negativ lutning är  $r = -1$ . Om ingen korrelation finns alls är  $r = 0$ . Se exemplen nedan.



Stark positiv korrelation.  
Ex. Fotlängd/skostorlek  
 $r \approx 1$



Svag negativ korrelation.  
Ex. Vägbredd/antal olyckor  
 $r \approx -0,5$



Ingen korrelation.  
Ex. Äggkonsumtion/  
kriminalitet.  $r \approx 0$

Ex. Medeltemperaturen och ett företags försäljning av glass under en vecka visas nedan.

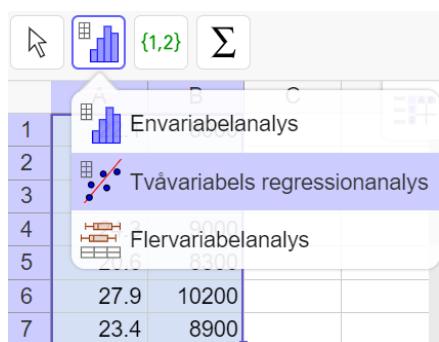
Medeltemperatur ( $^{\circ}\text{C}$ )	Glassförsäljning (kr)
22,1	8 000
18,5	7 500
26,7	9 500
24,3	9 000
20,6	8 300
27,9	10 200
23,4	8 900

- a) Rita ett spridningsdiagram över resultatet. Låt medeltemperaturen vara på x-axeln och glassförsäljningen på y-axeln.  
 b) Bestäm ekvationen för regressionslinjen.  
 c) Bestäm korrelationskoefficienten.  
 d) Uppskatta glassförsäljningen en vecka där medeltemperaturen är 30 °C utifrån din modell.

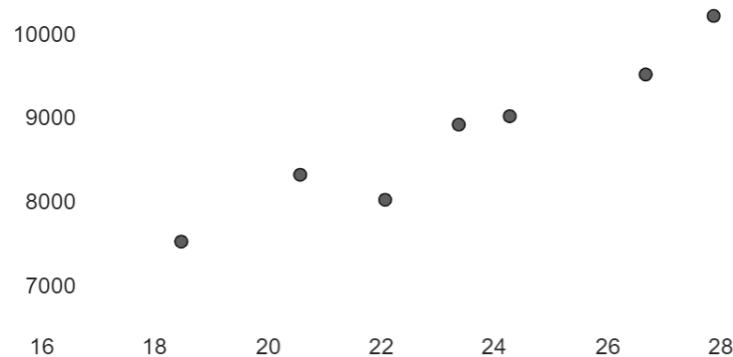
a) Starta Geogebra, växla till kalkylblad och skriv av värdena från tabellen.

	A	B
1	22.1	8000
2	18.5	7500
3	26.7	9500
4	24.3	9000
5	20.6	8300
6	27.9	10200
7	23.4	8900

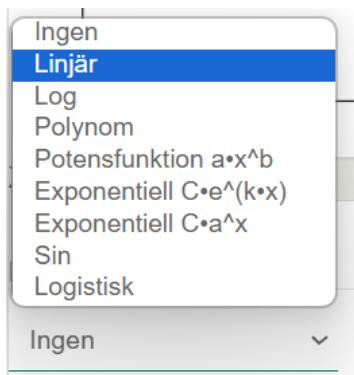
Markera värdena, klicka på "blå staplarna", välj *Tvåvariabels regressionsanalys* och vårt spridningsdiagram visas.



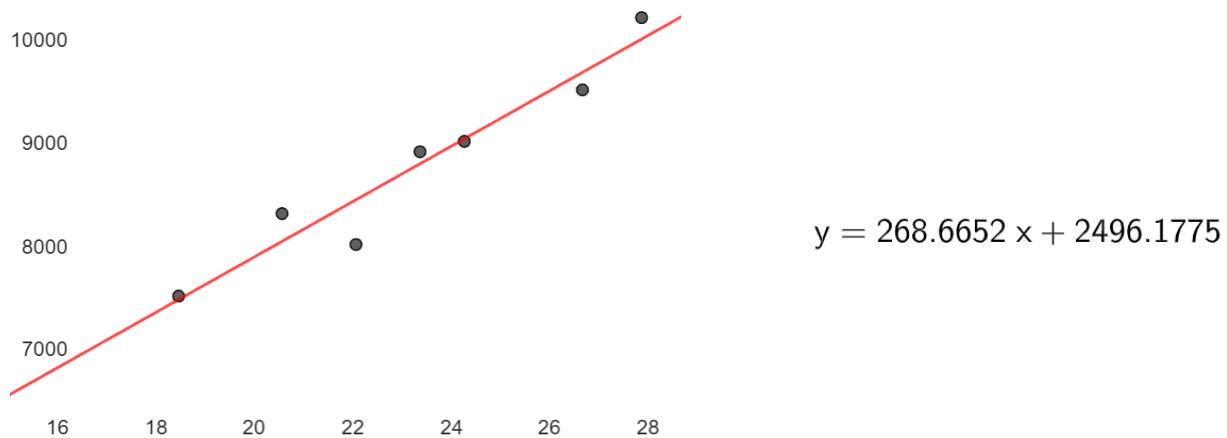
**Svar:**



b) Efter att ha tagit fram spridningsdiagrammet klickar vi på rullistan "Regressionsmodell" och väljer "Linjär".



Regressionslinjen ritas upp och ekvationen för den visas under diagrammet.



Vi avrundar värdena i ekvationen till lämpligt antal värdessiffer.

**Svar:**  $y \approx 270x + 2500$

(Givetvis går det bra att rita spridningsdiagrammet, rita regressionslinjen och ta fram ekvationen för hand, men det tar mycket längre tid och blir mindre exakt.)

c) Efter att ha tagit fram spridningsdiagrammet klickar vi på "Visa statistik"  $\Sigma x$ , läser av  $r$  och avrundar till två decimaler.

$S_x$	3.3035
$S_y$	919.6273
$r$	0.9651
$p$	0.9643
$S_{xx}$	65.4771

**Svar:**  $r \approx 0.97$

d) Från b-uppgiften fick vi fram sambandet  $y = 270x + 2\ 500$  där  $x$  är medeltemperaturen i °C och  $y$  försäljningen i kronor.

Om medeltemperaturen är 30 °C sätter vi in  $x = 30$  i formeln ovan för att få fram den förväntade försäljningen.

$$y = 270 \cdot 30 + 2\ 500 = 10\ 600$$

**Svar:** 10 600 kr