



AILIT



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



Svenska
kulturfonden

Från medelvärde till maskininläring



Maskininlärning bygger på data

Maskininlärning använder **data** för att skapa en modell som kan beskriva verkligheten. **Modellen** kan till exempel användas för att förutspå något om situationer med liknande data.

En modell är alltså en förenklad version av något som finns i verkligheten – den tar med det som är viktigt och lämnar bort resten.

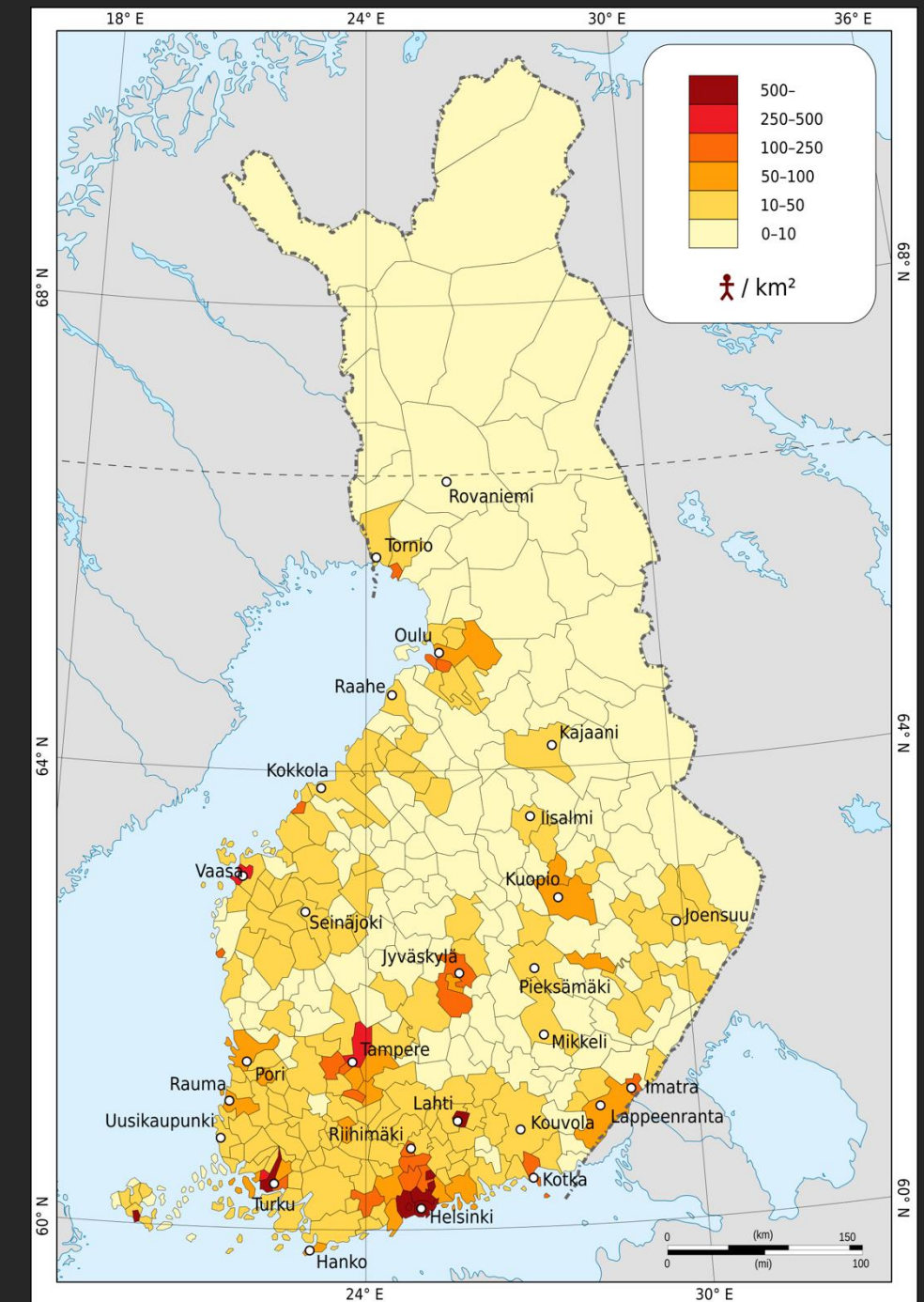
Vi kommer i den här lektionen att fundera på glassförsäljning!



Repetition: exempel på modeller

Tänk er som exempel en karta av Finland jämfört med verkligheten!

Kartan visar inte varje sten eller buske, men den visar det vi behöver för att hitta rätt.



Repetition: exempel på modeller

En modell kan också vara matematisk, och beskriva ett samband istället för en plats.

Om vi vet att en person cyklar 15 km/h är detta en enkel modell av sambandet mellan hastighet och tid, som vi till exempel kan använda för att räkna ut hur lång tid det tar för personen att cykla 100 kilometer.

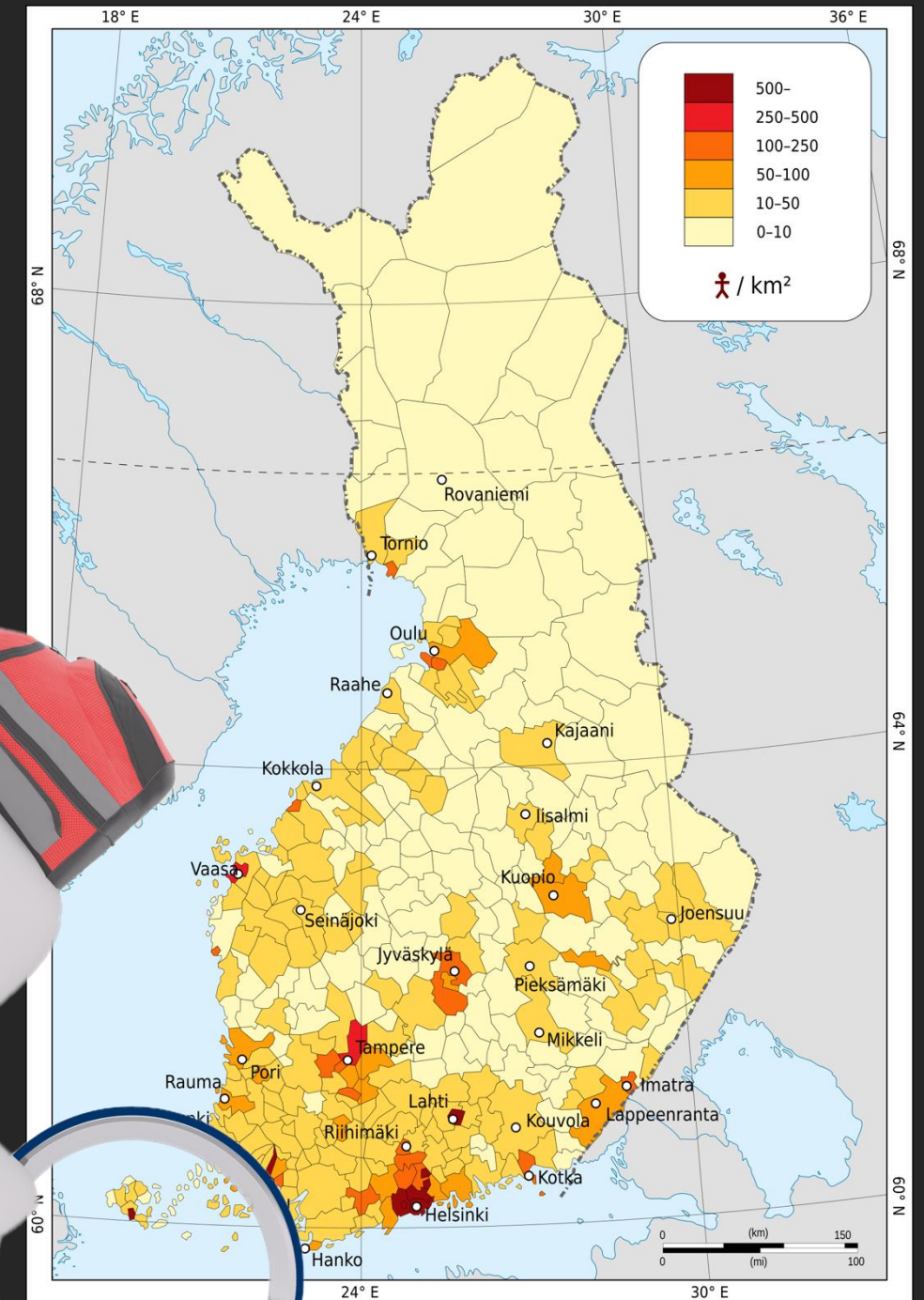
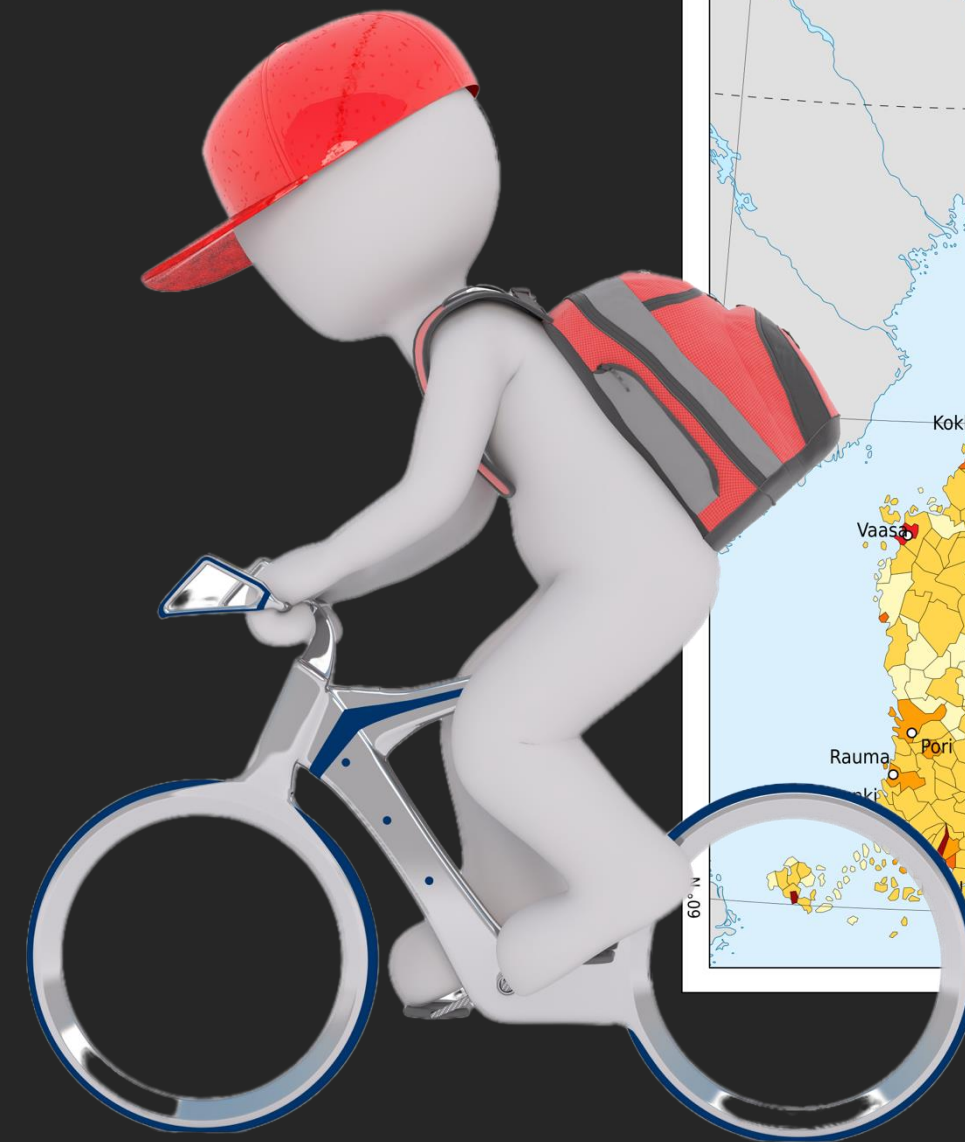
Precis som kartan inte tar hänsyn till varje detalj i landskapet, tar den här modellen inte hänsyn till att cyklisten kanske blir trött eller möter motvind – den fångar bara det viktigaste sambandet.



Repetition: exempel på modeller

Kartan och modellen för att beräkna sambandet mellan hastighet och tid har vi människor tagit fram.

Med hjälp av AI och maskininlärning kan vi istället låta datorn hitta liknande samband genom att analysera stora mängder data.



Vad vi ska göra...



Glasskiosken Strutkungen vill få bättre koll på sin försäljning. Vi ska bygga en matematisk modell som förutspår hur många glassar de säljer. Samtidigt lär vi oss om hur AI-modeller fungerar.

1

Börja enkelt

Räkna ut ett medelvärde, det blir vår första modell.

2

Hitta mönstret

Lägg till mer information och se vad som händer.

3

Bygg en ny modell

Rita en linje och tolka vad k och m betyder.

4

Möt verkligheten

Justera modellen i ett verktyg.



Glasskiosken Strutkungen



Hur många glassar ska vi förbereda till imorgon?

Om vi förbereder för få

→ förlorad försäljning, missnöjda kunder

Om vi förbereder för många

→ glass slängs, pengar förloras

10 dagars försäljning

Glasskioskens ägare håller koll på hur många glassar som säljs per dag, och de har gett oss information om de 10 senaste dagarna.

Vilken typ av modell kan vi skapa utifrån 10 dagars försäljningsdata?

Medelvärde kan vara en bra modell!

Dag	Sålda glassar
1	60
2	110
3	35
4	75
5	25
6	90
7	70
8	110
9	45
10	80

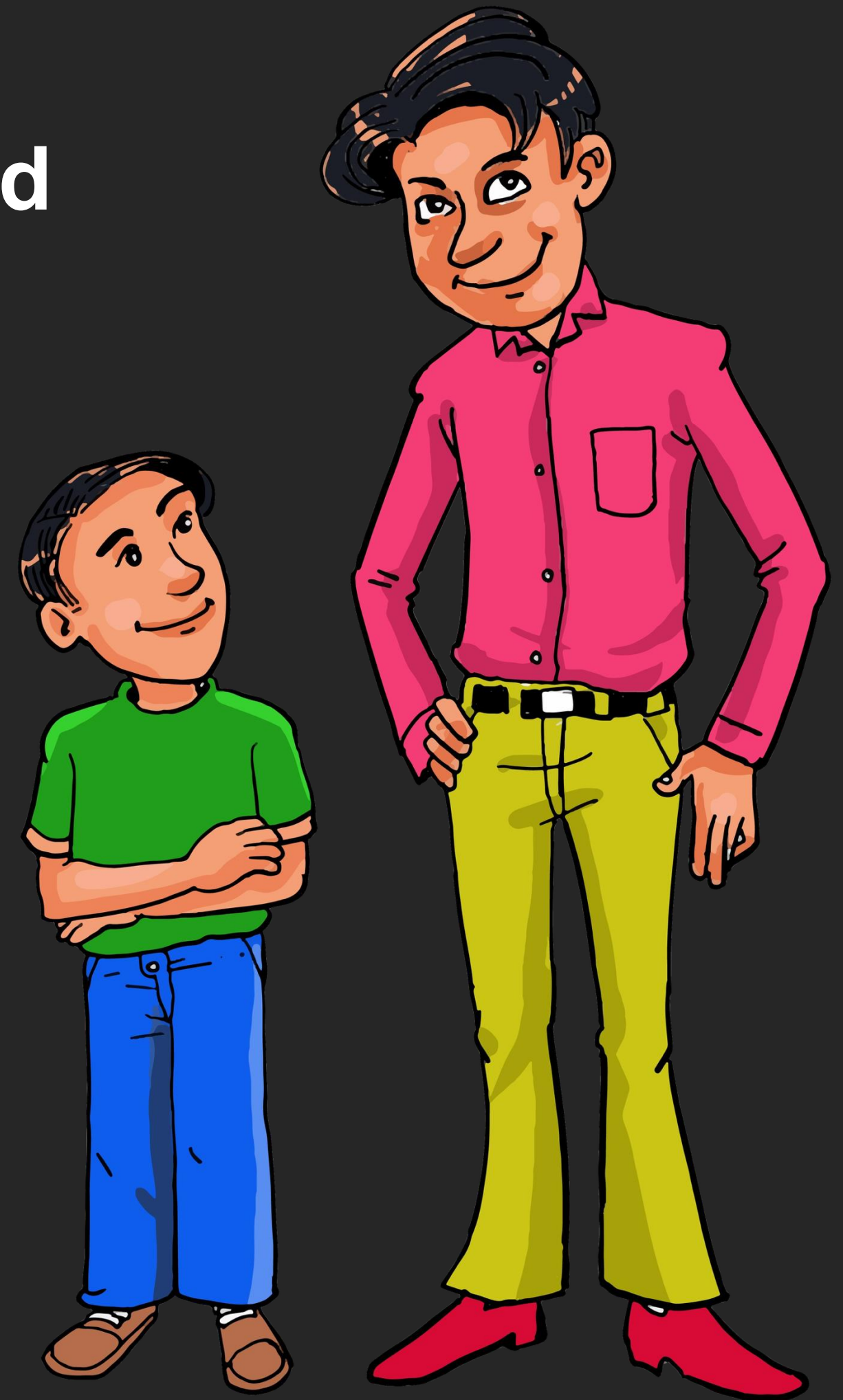
[Valfritt material]

Exempel: medellängd

Medellängden för män i Finland 181 cm.

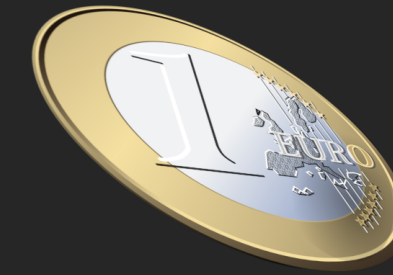
Det betyder att en slumpmässigt vald man troligen kommer vara nära 181 cm, men variationen mellan olika män är stor.

En del är kortare, andra längre.



[Valfritt material]

Exempel: singla slant



Det samma gäller när vi singlar slant. Vi kan få följande resultat om vi singlar den fem gånger:



Eller på motsvarande sätt 5 x klave.

Den bästa gissningen är ändå att fördelningen kommer vara nära 50/50: om vi kastar tillräckligt många gånger kommer vi få krona ungefär hälften av gångerna, och klave ungefär hälften av gångerna.



Beräkna medelantalet glassar

UPPGIFT

Räkna ut medelvärde

1. Addera alla 10 dagars försäljning.
2. Dela summan med antalet dagar (10).
3. Det här är vår *första modell*: en gissning på hur många glassar som säljs en typisk dag.

$$\text{Medelvärde} = \frac{\text{summan av alla värden}}{\text{antalet värden}}$$

Dag	Sålda glassar
1	60
2	110
3	35
4	75
5	25
6	90
7	70
8	110
9	45
10	80

Beräkna medelantalet glassar

M E D E L V Ä R D E T

70

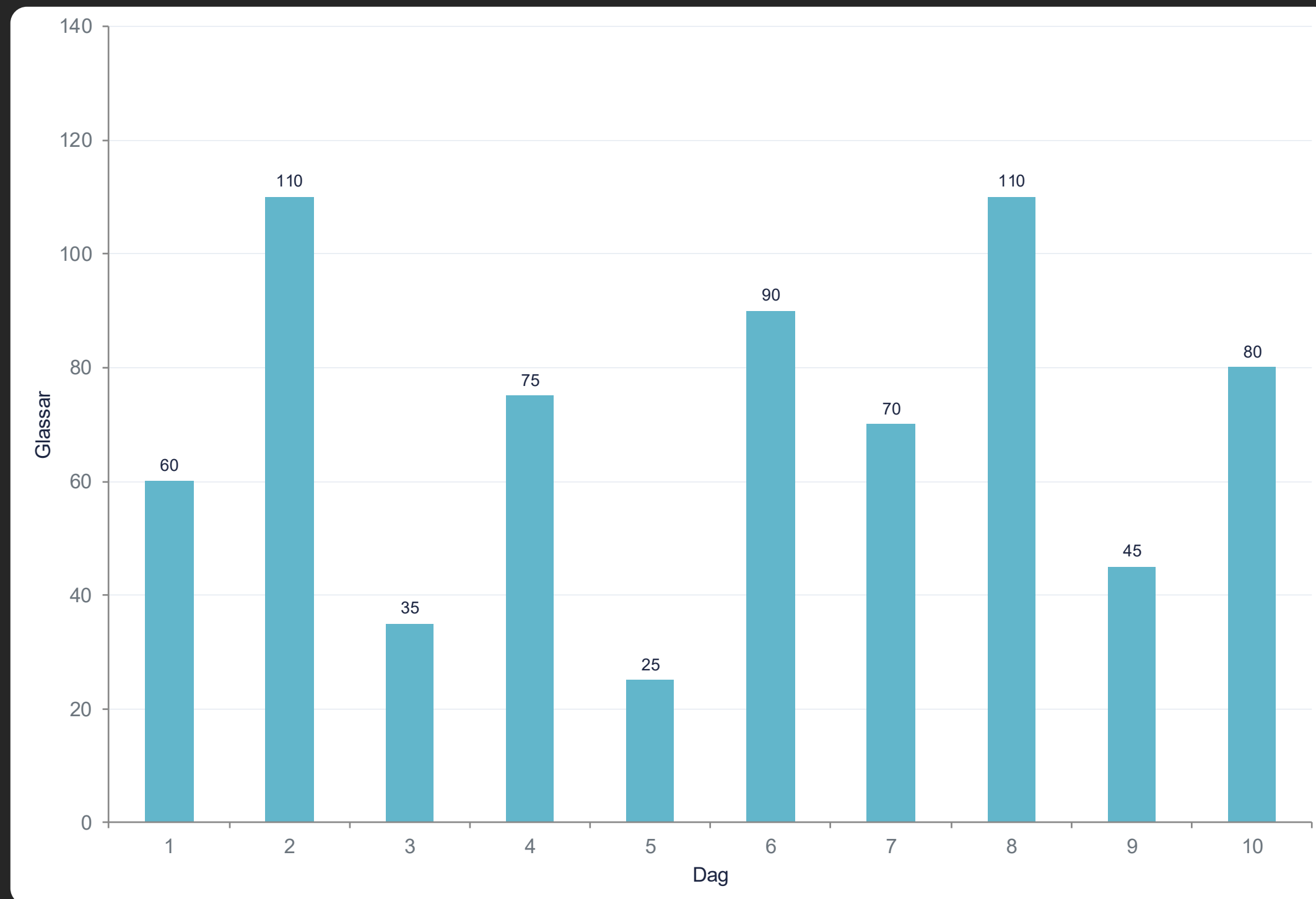
glassar per dag

Vår första modell:

*Vi gissar att vi
behöver förbereda
70 glassar per dag.*

Dag	Sålda glassar
1	60
2	110
3	35
4	75
5	25
6	90
7	70
8	110
9	45
10	80

Är 70 en bra gissning?



Här är samma värden presenterade i ett diagram.

Vad kan vi säga utifrån det?

- Vilka dagar var medelvärdet långt ifrån verkligheten?
- Vad kan förklara skillnaderna?
- Vilken information saknar vi?

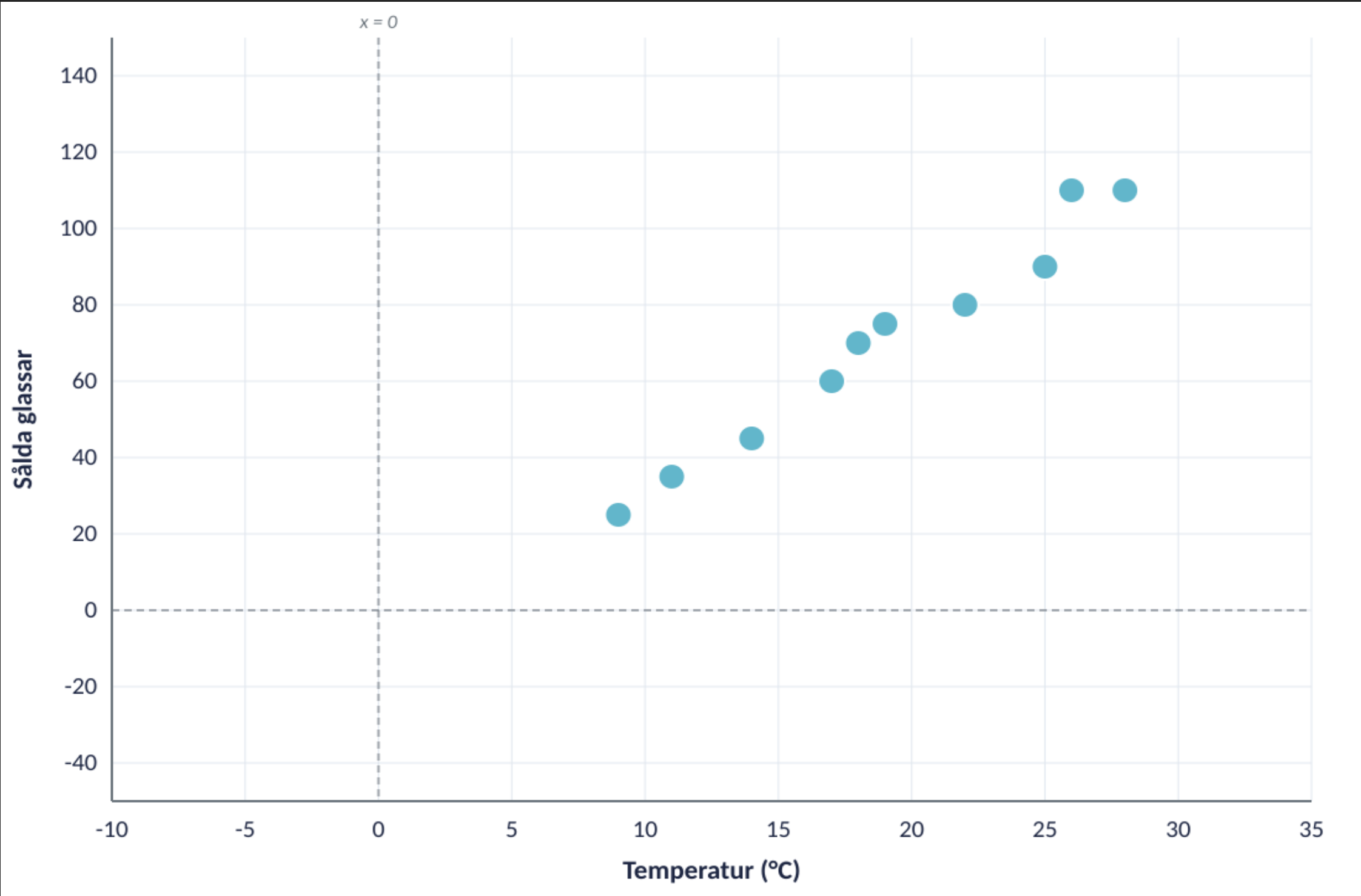


Lägg till temperaturen

Ny information som gör att vi kan tolka försäljningsdata på ett nytt sätt.

Samma försäljningsdata – nu med temperatur

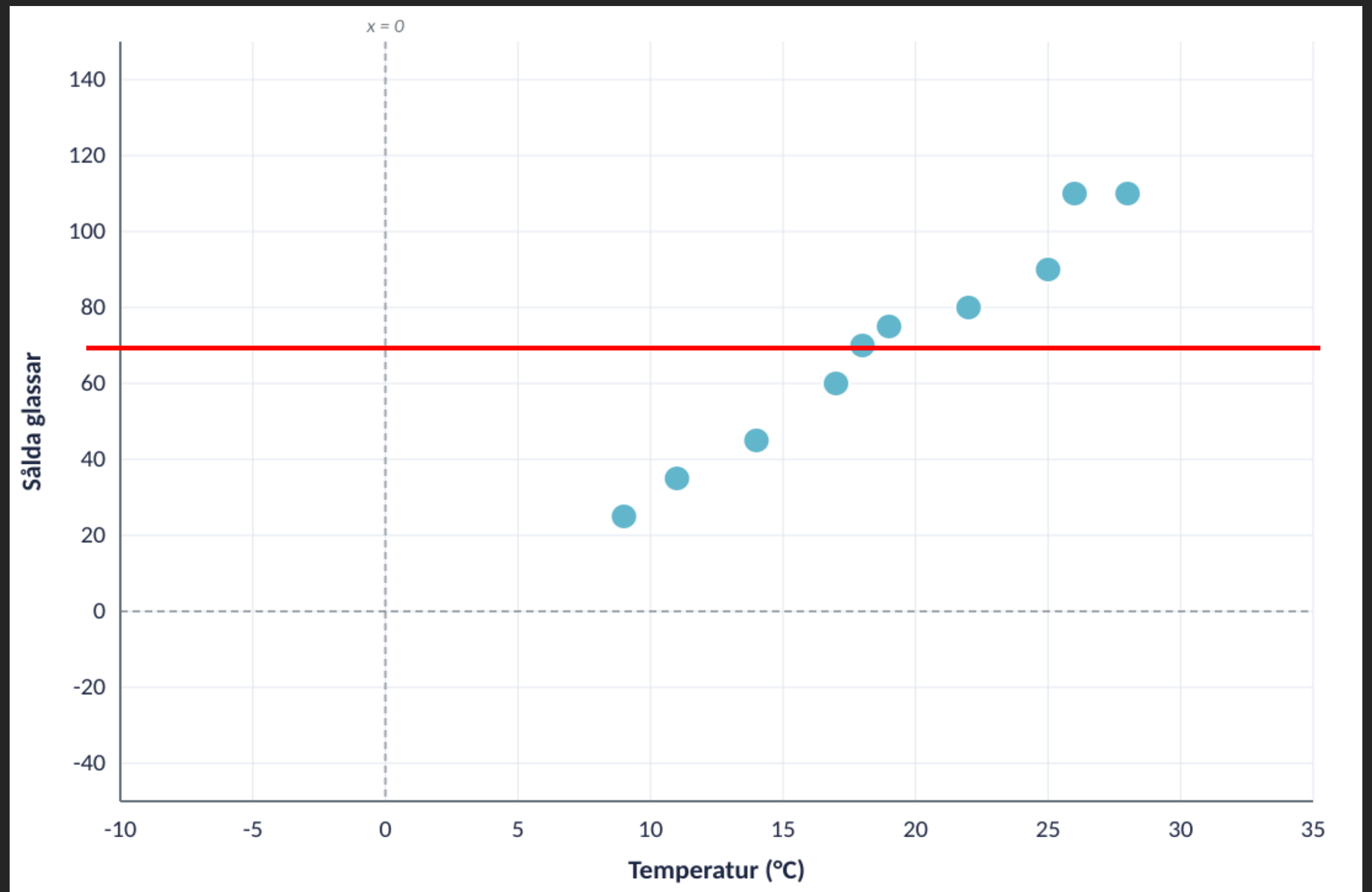
Dag	Temp °C	Sålda glassar
1	17	60
2	26	110
3	11	35
4	19	75
5	9	25
6	25	90
7	18	70
8	28	110
9	14	45
10	22	80



Nu ser vi hur fel dålig vår första modell (medelvärdet) var

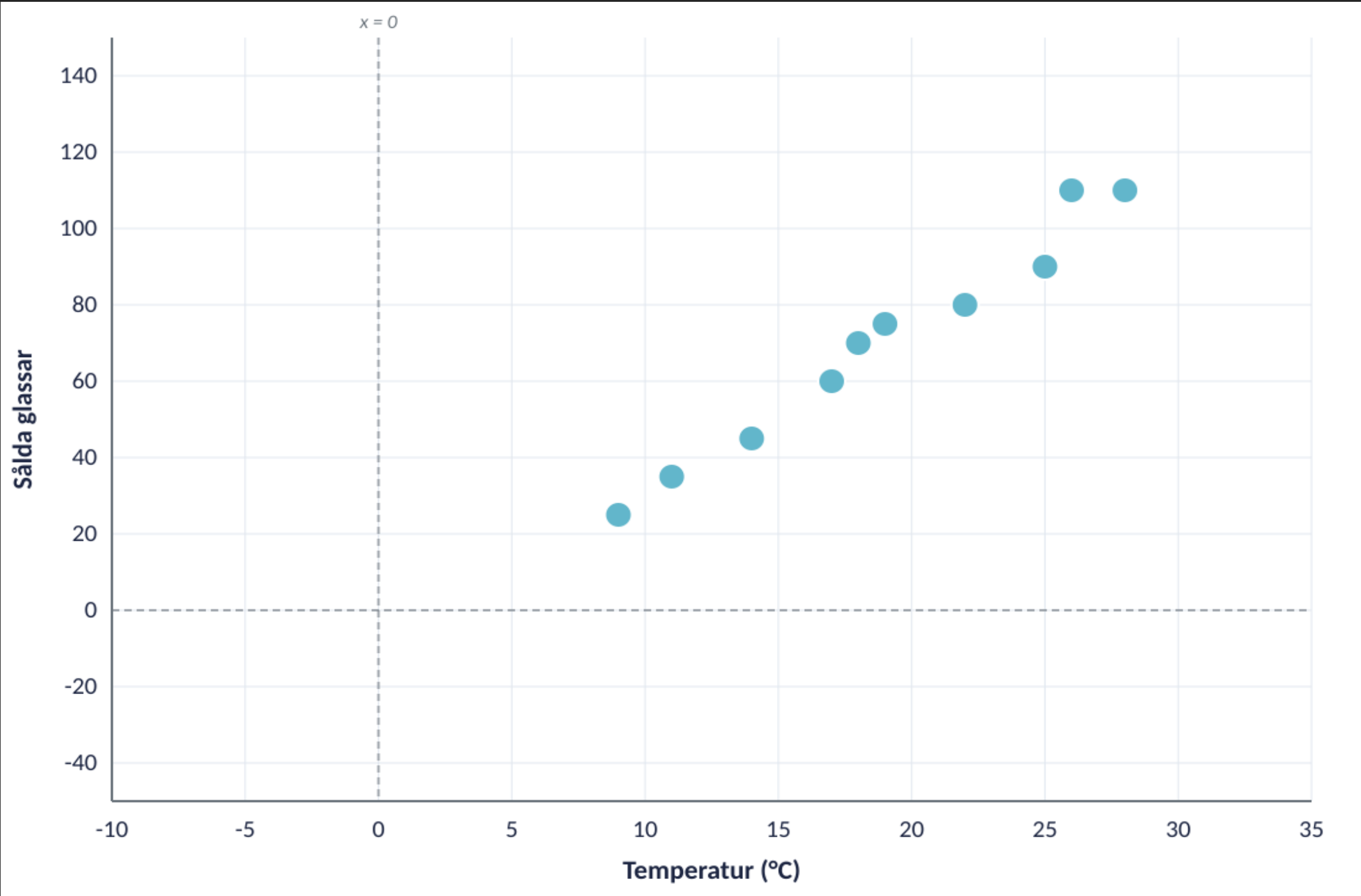
Dag	Temp °C	Sålda glassar
1	17	60
2	26	110
3	11	35
4	19	75
5	9	25
6	25	90
7	18	70
8	28	110
9	14	45
10	22	80

Medelvärdet ($y=70$) är inte en bra modell för hur många glassar vi behöver förbereda.



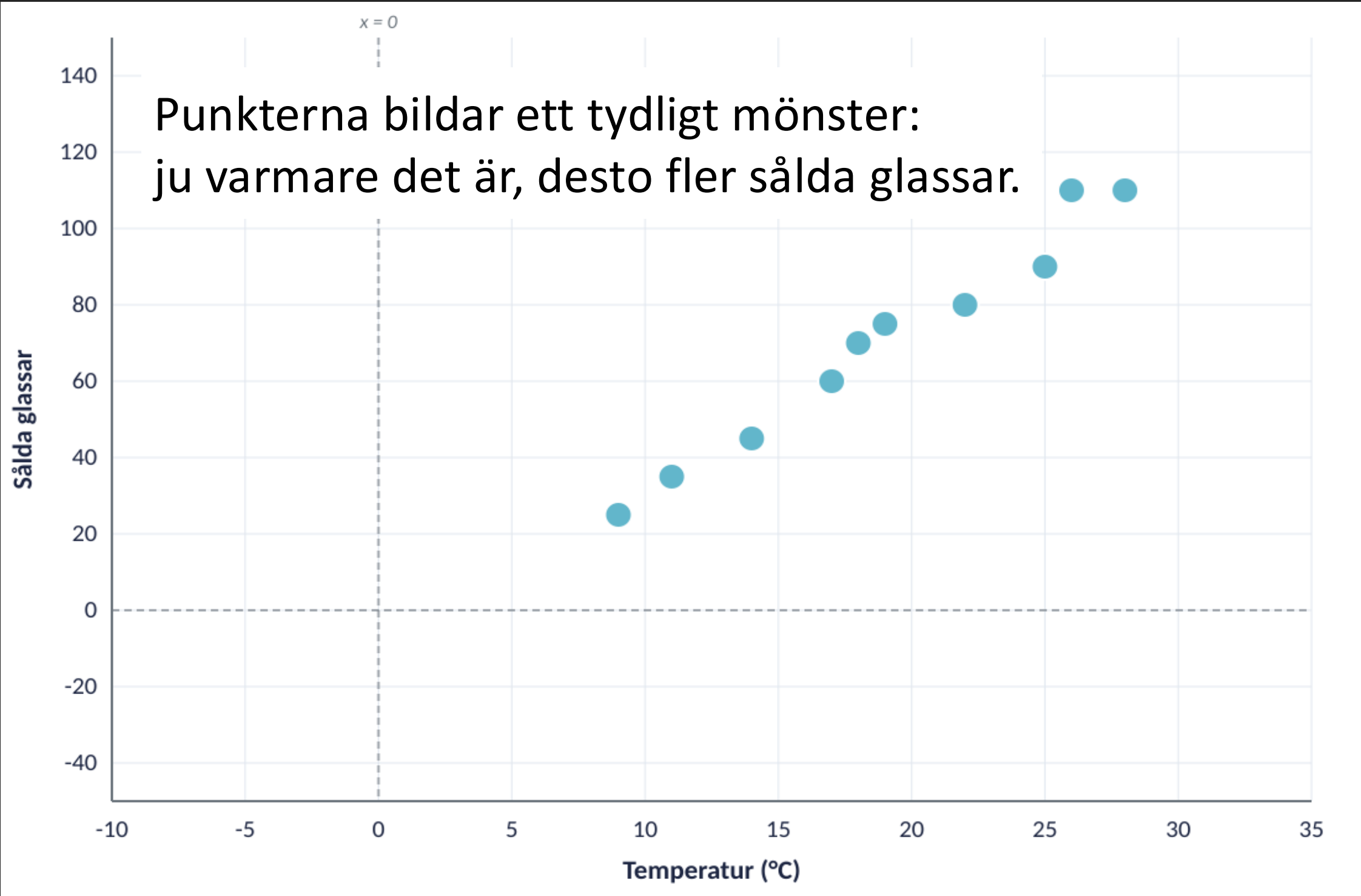
Vad visar diagrammet?

Dag	Temp °C	Sålda glassar
1	17	60
2	26	110
3	11	35
4	19	75
5	9	25
6	25	90
7	18	70
8	28	110
9	14	45
10	22	80



Vad visar diagrammet?

Dag	Temp °C	Sålda glassar
1	17	60
2	26	110
3	11	35
4	19	75
5	9	25
6	25	90
7	18	70
8	28	110
9	14	45
10	22	80





Bygg en linjär modell

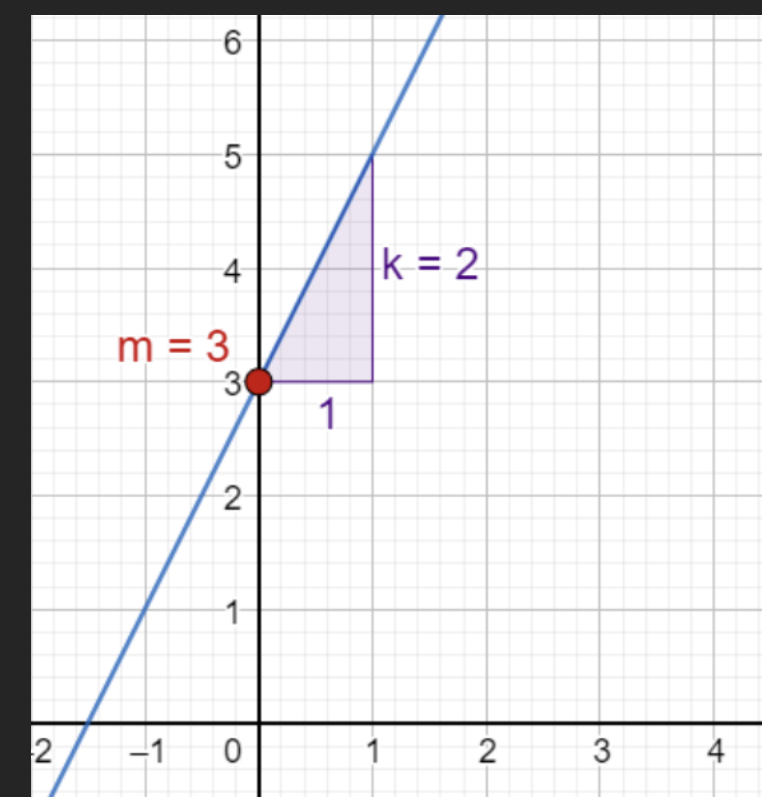
lutning / riktningskoefficient



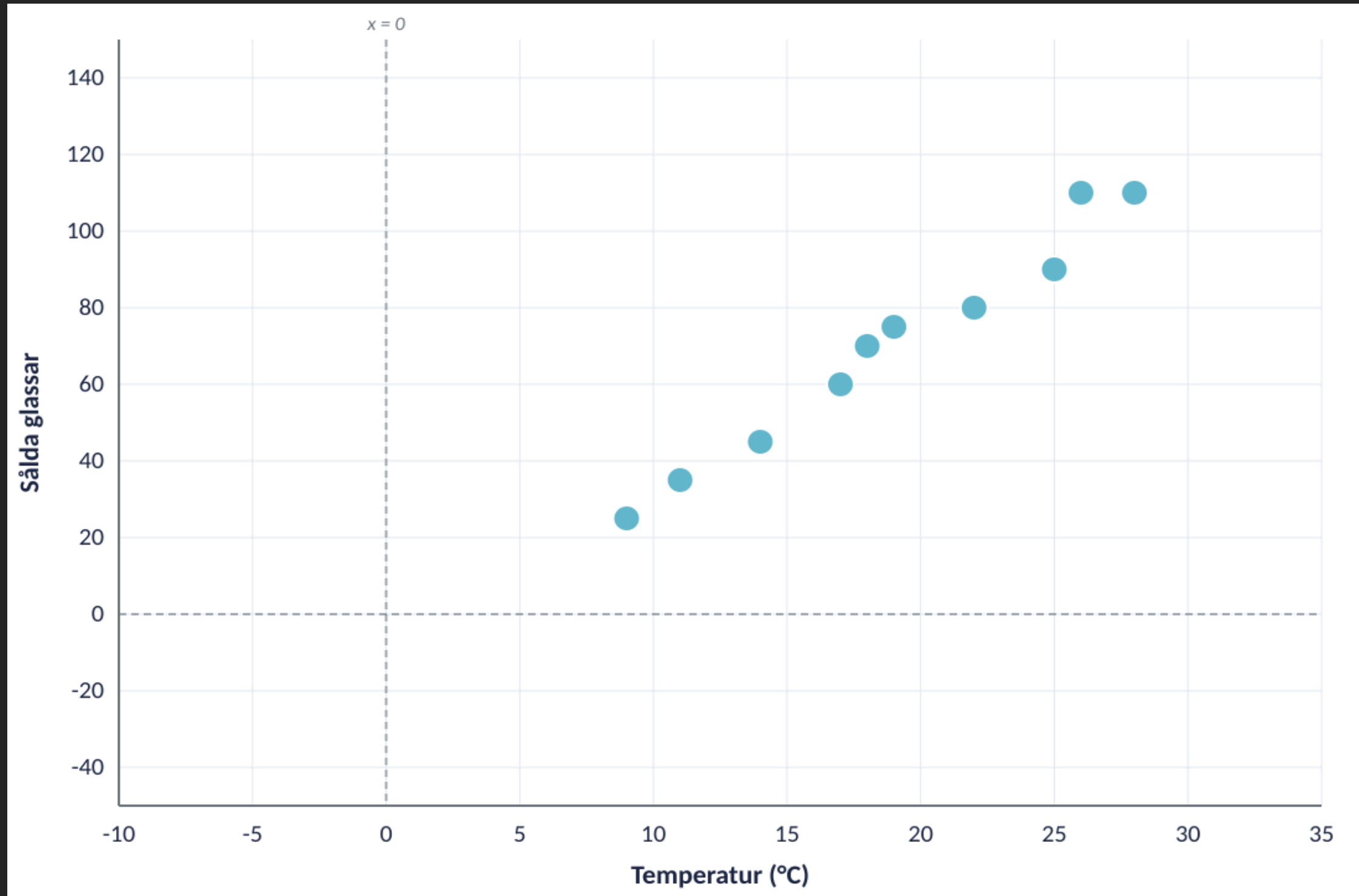
Vi kan använda den räta linjens ekvation: $y = kx + m$.



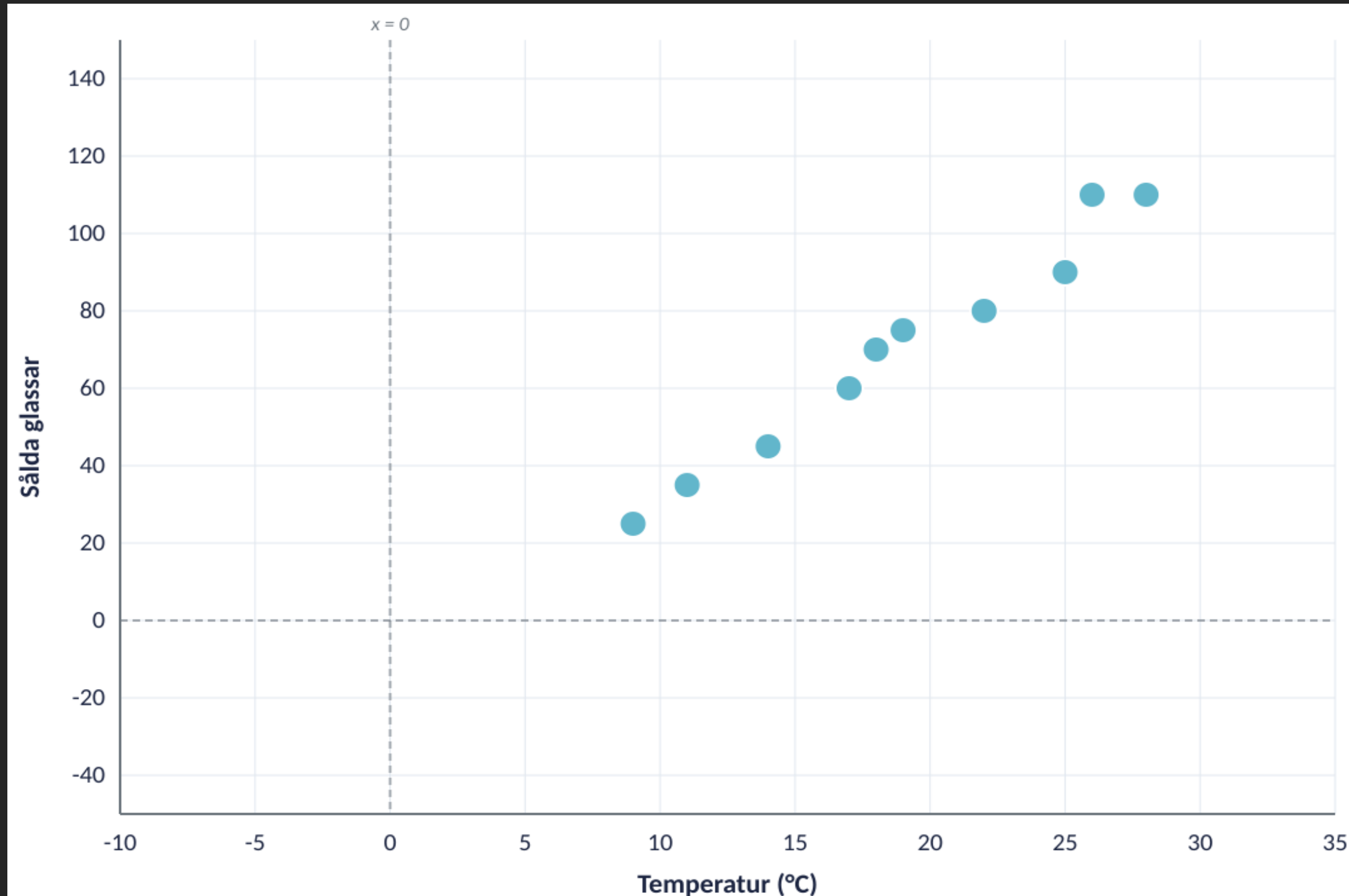
linjens skärningspunkt på y-axeln



Kan vi förutsäga glassförsäljningen för olika temperaturer genom att dra en rät linje genom punkterna?

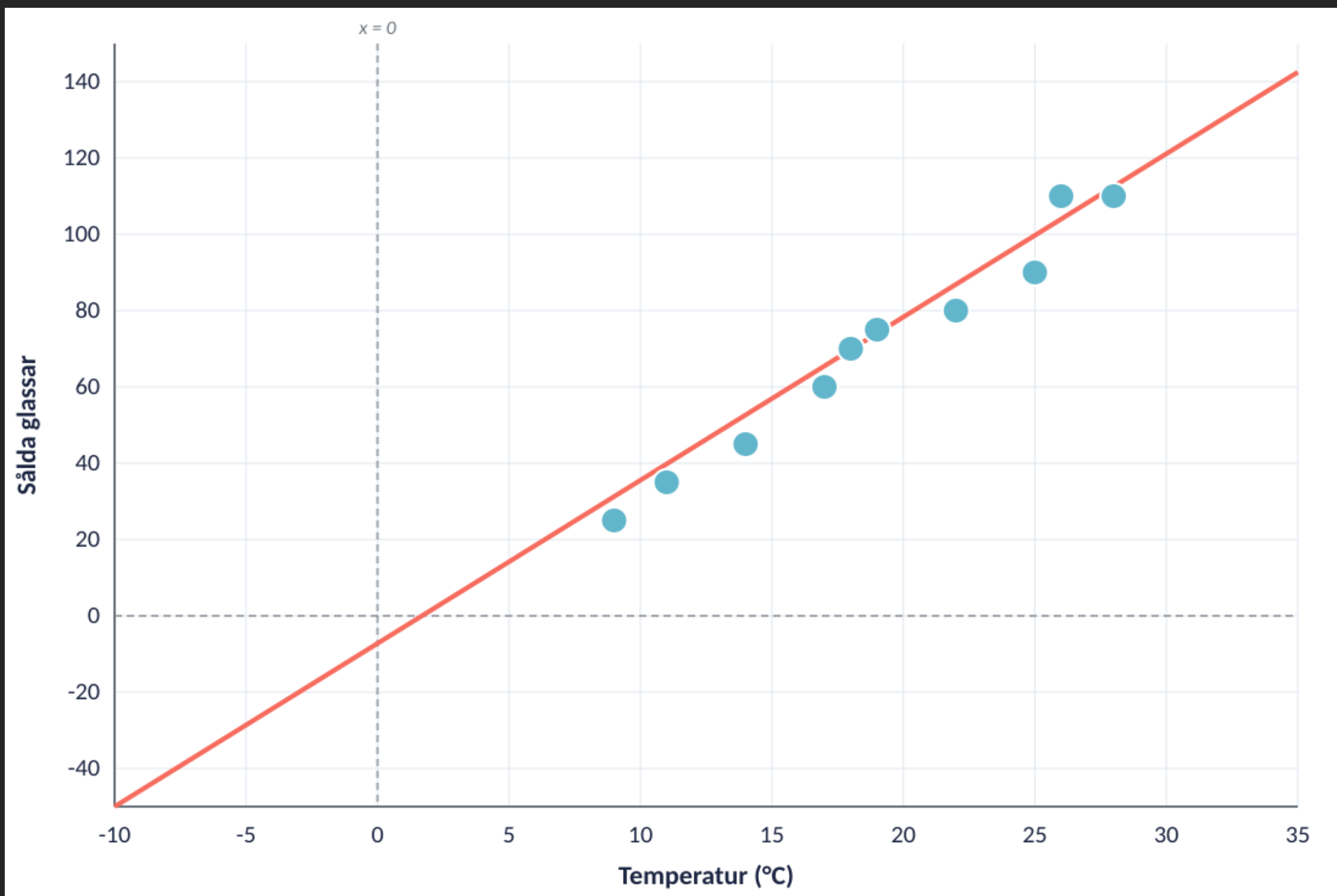


Nej, men vi kan dra en linje som är nära alla punkter



1. Rita ut punkterna i ett koordinatsystem på papper.
2. Dra en **rak linje** med linjal som passar punkterna så bra som möjligt.
3. Lyckas du uppskatta ***k*** (lutningen)?
4. Lyckas du uppskatta ***m*** (skärningspunkten där linjen skär y-axeln)?

Kanske ser din linje ut så här?

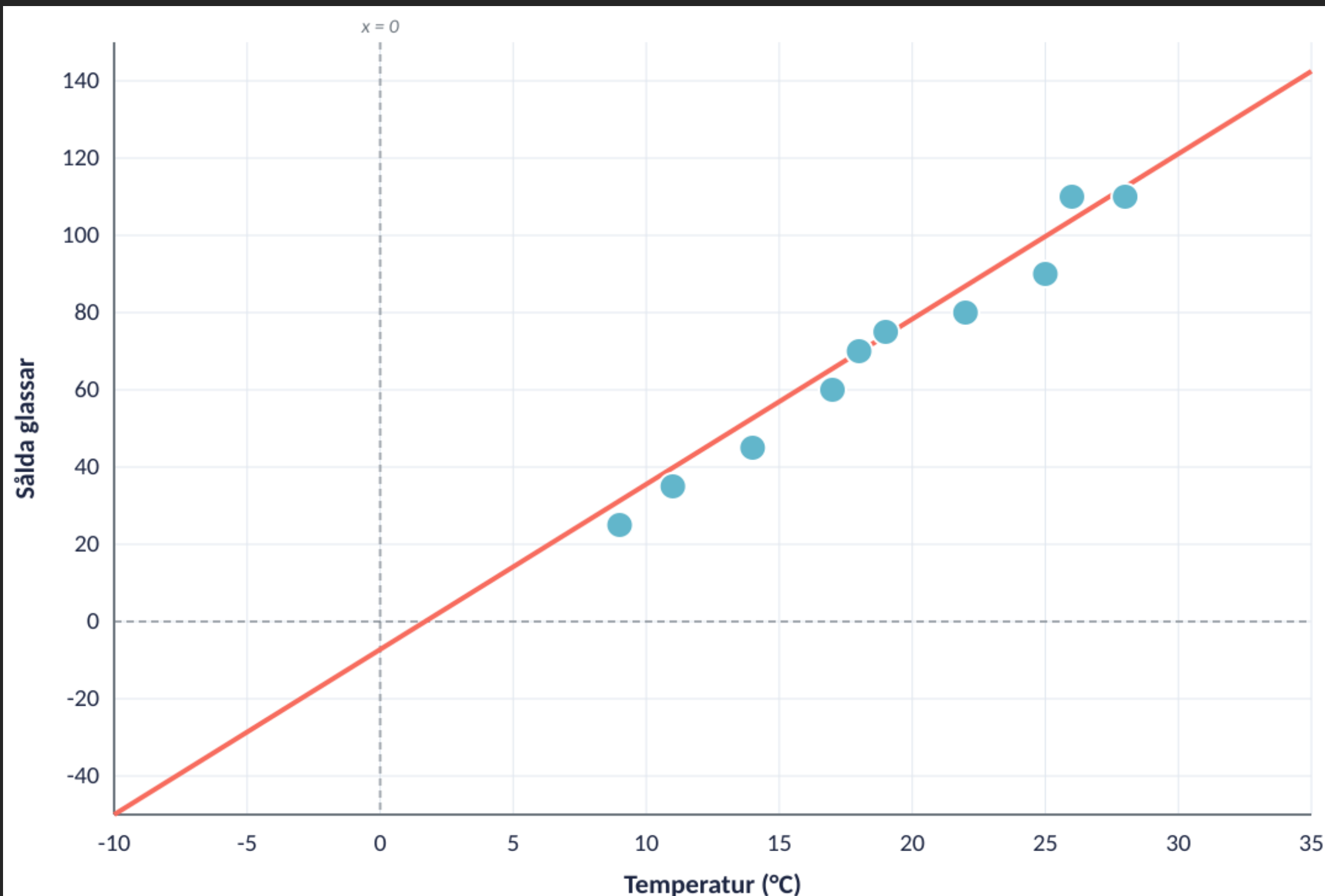


$$y \approx 4.5x - 15$$

Vår modell säger:

Vi säljer ungefär 4,5 fler glassar för varje grad varmare temperatur.

Kanske ser din linje ut så här?



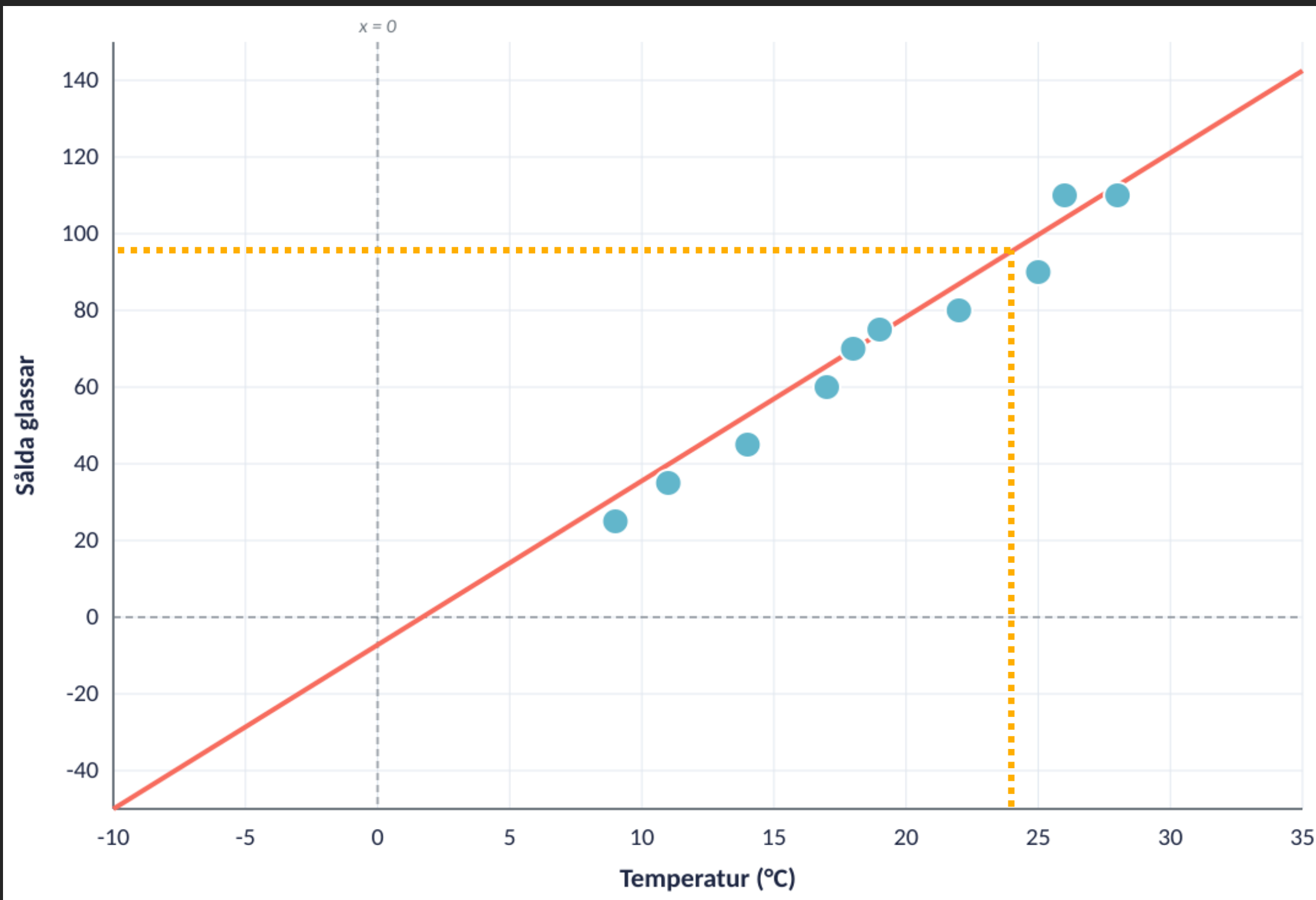
$$y \approx 4.5x - 15$$

Vår modell säger:

Vi säljer ungefär 4,5 fler glassar för varje grad varmare temperatur.

Hur många glassar behövs om det utlovas 24 grader imorgon?

Kanske ser din linje ut så här?



$$y \approx 4.5x - 15$$

Vår modell säger:

Vi säljer ungefär 4,5 fler glassar för varje grad varmare temperatur.

Hur många glassar behövs om det utlovas 24 grader imorgon?

$$y \approx 4,5 \cdot 24 - 15$$

≈ 93 glassar

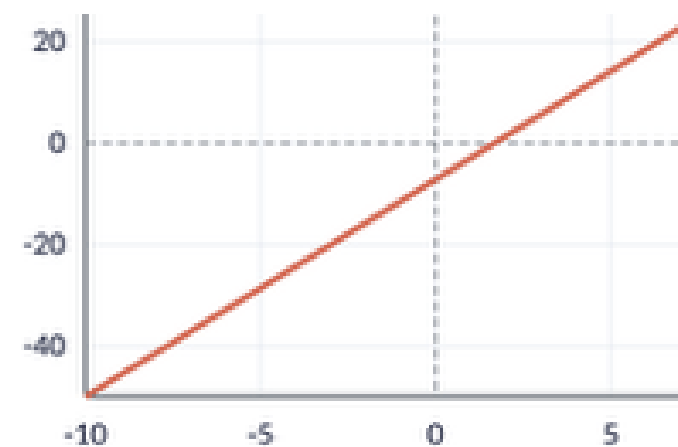
Vad innebär k och m i praktiken?

Vad innebär k och m i praktiken?

k

Lutningen

Hur mycket y ändras när x ökar med 1.



I vår modell:

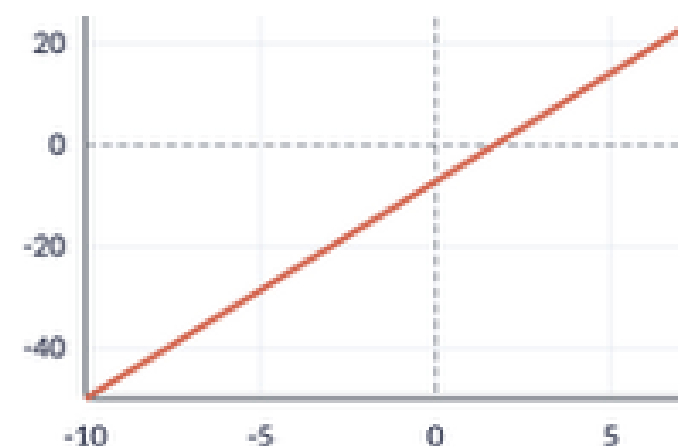
för varje grad varmare det blir,
säljs ungefär 4,5 fler glassar.

Vad innebär k och m i praktiken?

k

Lutningen

Hur mycket y ändras när x ökar med 1.



I vår modell:

för varje grad varmare det blir,
säljs ungefär 4,5 fler glassar.

m

Skärningspunkten

Var linjen skär y -axeln (när $x = 0$).

I vår modell:

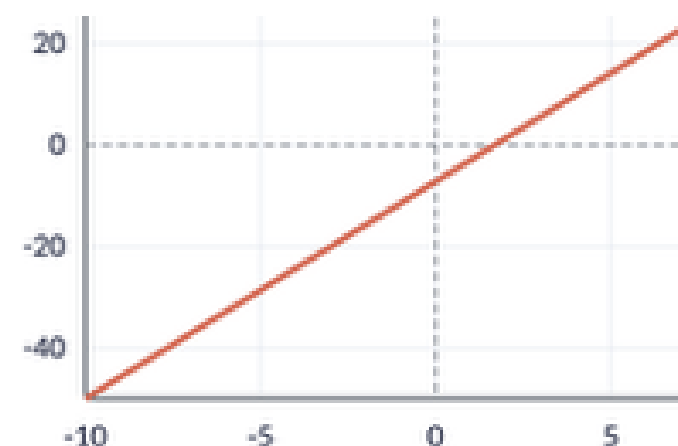
vid 0°C kommer vi att sälja -15 glassar.

Vad innebär k och m i praktiken?

k

Lutningen

Hur mycket y ändras när x ökar med 1.



I vår modell:

för varje grad varmare det blir,
säljs ungefär 4,5 fler glassar.

m

Skärningspunkten

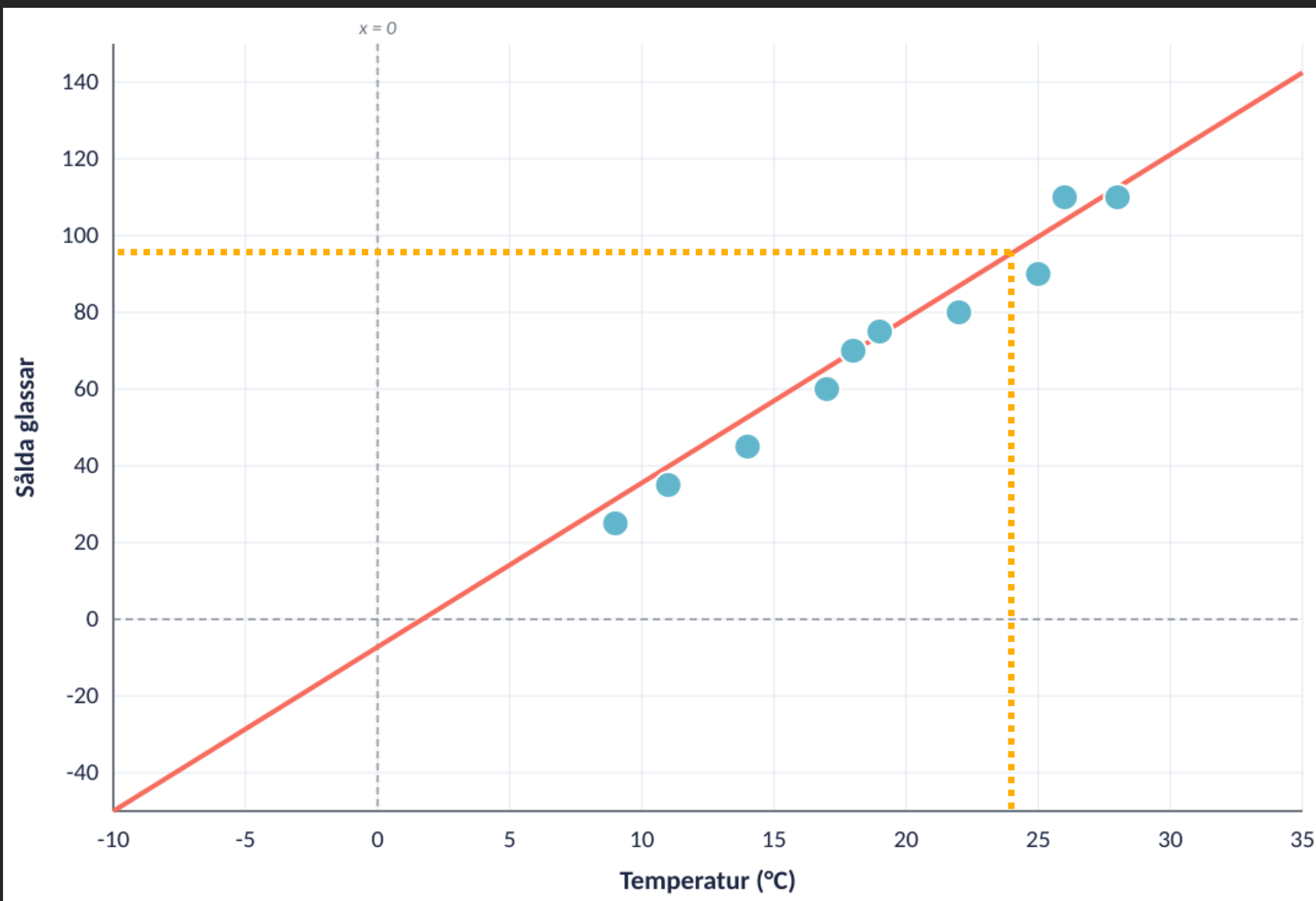
Var linjen skär y -axeln (när $x = 0$).

I vår modell:

vid 0°C kommer vi att sälja **-15 glassar.**

Är det rimligt?

Fel i modellen – ett negativt antal är inte rimligt



Varför blir det så här?

Data från verkliga livet följer så gott som aldrig en helt rät linje.

Det betyder att det blir ett avstånd mellan linjen och de olika data-punkterna. Det här avståndet kallas modellens **fel**. Ju större avstånd, desto större fel.

Alla felen tillsammans visar hur bra modellen stämmer med verkligheten, dvs. hur bra den passar våra data.

Maskininlärning justerar linjen så att felen minskar. Målet är att göra det totala felet så litet som möjligt.



Använd det interaktiva verktyget

Lägg in egna data, justera skärningspunkten och lutningen.

A

Drömvärlden

Perfekta data

B

Verkligheten

Inte lika perfekta data

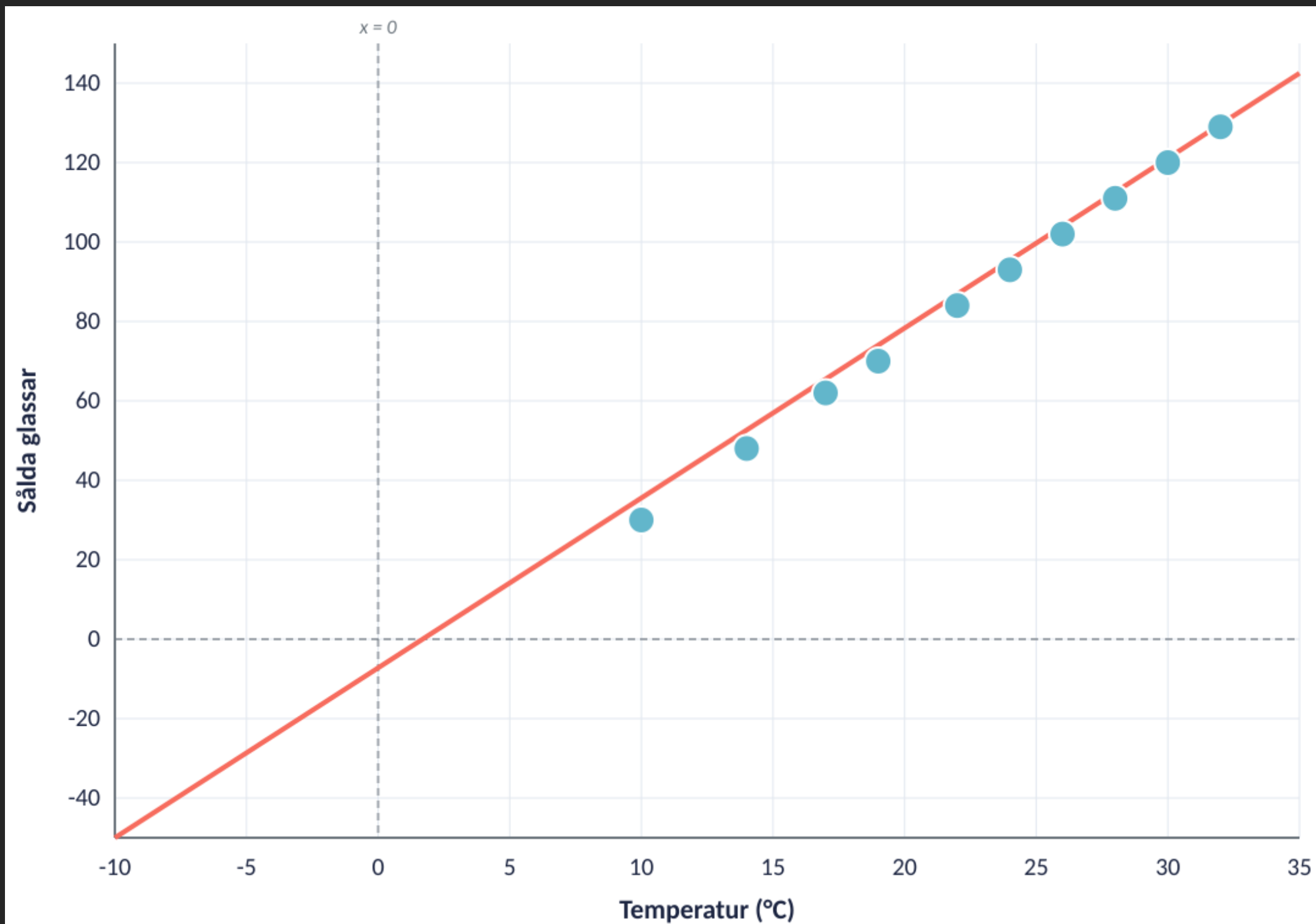
C

Nya dagar

Modellen uppdateras

A

Drömvärlden



Ändra lutningen och skärningspunkten tills linjen passar punkterna så bra som möjligt.

- Hur ändras linjen när du ändrar lutning och skärningspunkt?
- Hur bra blir modellen?

UPPGIFT

1 Vilket k-värde och m-värde fick du?

2 Vad betyder ditt k-värde? För varje grad varmare det blir, säljs ungefär ____ fler glassar.

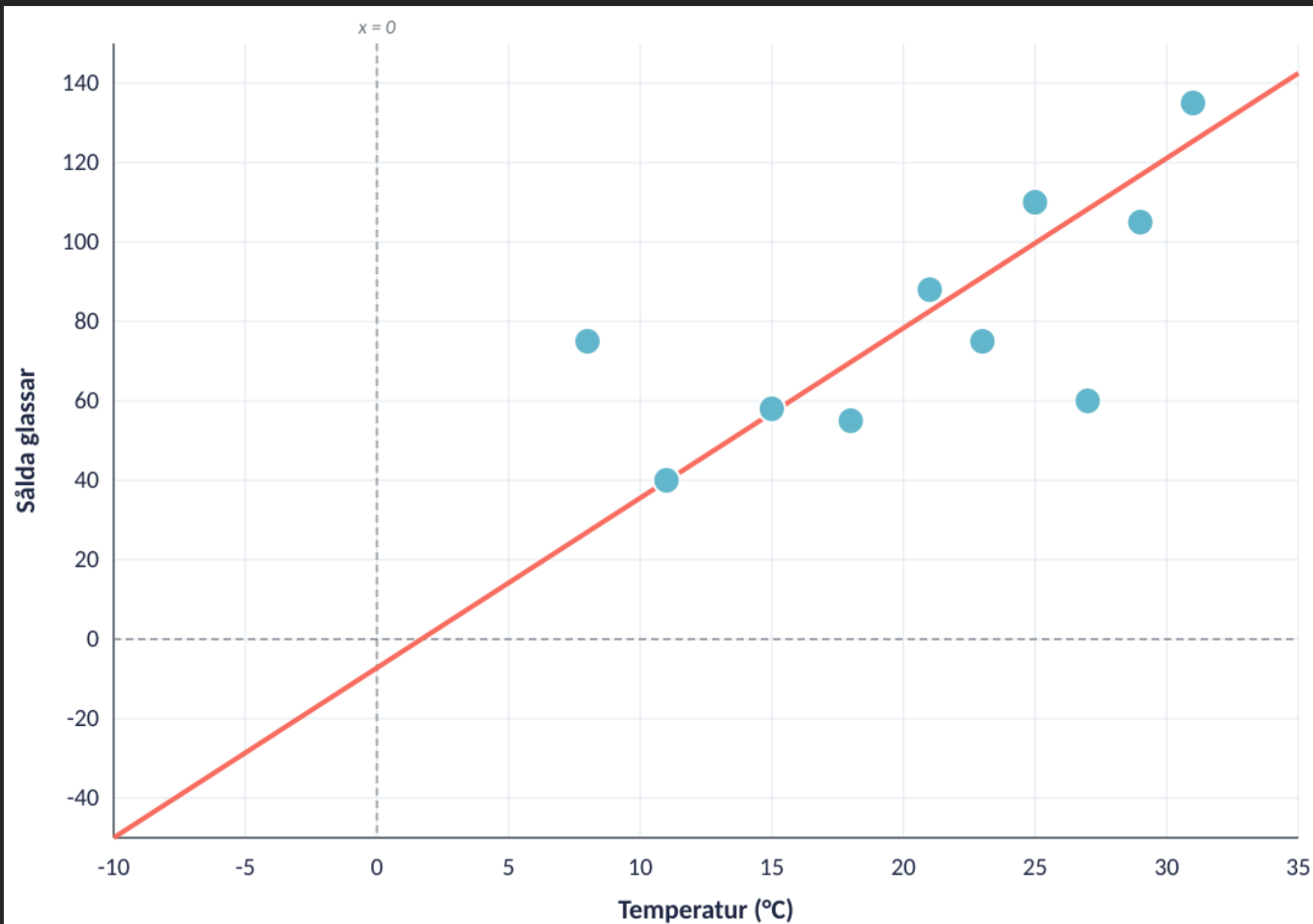
3 Vad betyder m-värdet? Vad säger din modell om antalet glassar när det är 0°C ?

4 Imorgon ska det bli 23°C . Hur många glassar bör Strutkungen förbereda?

5 Vad händer om du sätter $k = 0$? Hur ser linjen ut då?

B

Verkligheten



Data följer sällan en linje så bra som ovan. Om de gjorde det skulle vi aldrig behöva göra en modell för att uppskatta något, eftersom den skulle vara exakt. Men vissa äter 10 glassar när det är -20 och ingen alls när det är varmt.

Felmarginal

Verktyget visar hur långt punkterna är från linjen.

--> Mål: att hitta det minsta möjliga totalfel.

B Data

Dag	Temperatur (°C)	Sålda glassar
1	11	40
2	15	58
3	18	55
4	27	60
5	21	88
6	23	75
7	25	110
8	29	105
9	8	75
10	31	135

EXTRA

Om du t.ex. vill ge eleverna data att jobba med utanför verktyget. Detta är inte en del av själva presentationsmaterialet.

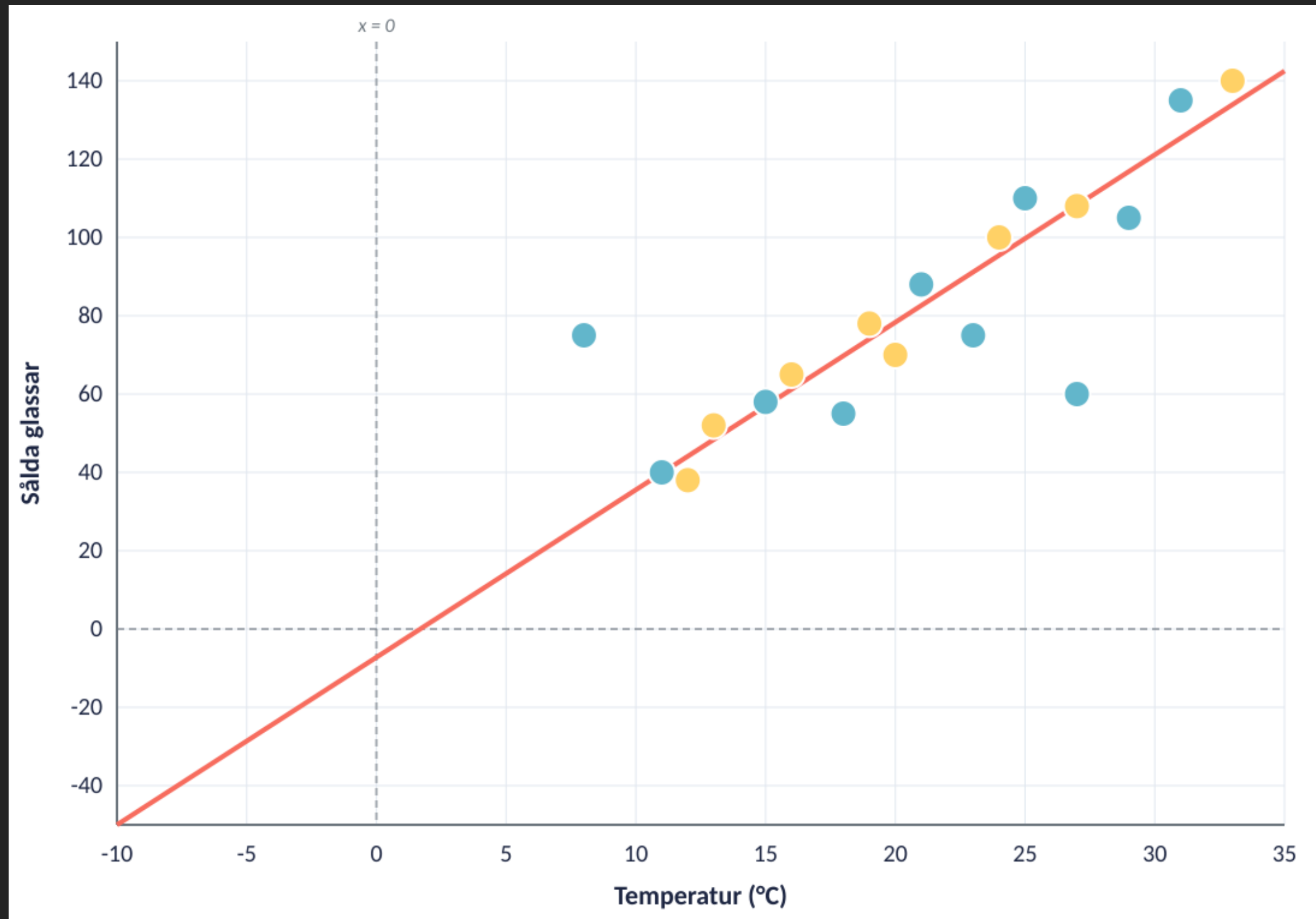
UPPGIFT

6 Försök hitta den linje som passar bäst totalt sett. Hur litet kan du få medelfelet?

7 Skriv ner ditt nya k och m . Skiljer de sig mycket från drömvärlden?

8 Ge minst en realistisk förklaring till varför en av de udda punkterna ligger där den gör.

c Nya dagar



● Tidigare data ● Data från nya dagar

Vi behöver kunna uppdatera modellen när vi får nya data.

Maskininlärning handlar om att träna modellen, testa, och träna igen.

Data

Dag	Temperatur (°C)	Sålda glassar
11	13	52
12	16	65
13	20	70
14	24	100
15	27	108
16	19	78
17	33	140
18	12	38

EXTRA

Om du t.ex. vill ge eleverna data att jobba med utanför verktyget. Detta är inte en del av själva presentationsmaterialet.

UPPGIFT

9 Vad händer med medelfelet när nya punkter kommer in? Ökar eller minskar det?

10 Justera linjen så att den passar alla punkter så bra som möjligt. Hur mycket ändrade du k och m ?

Matematiken bakom AI

Matematik driver AI-modellerna.



Räkna ut den bästa riktningen

För två variabler (glass och temperatur) lyckades vi hitta en tillräckligt bra och enkel formel.



Med fler variabler

Med många variabler blir det för komplicerat att räkna för hand.



Datorn räknar åt oss

Datorerna utför de beräkningar som behövs mycket snabbare.



När antalet variabler växer...

Då kan vi inte längre hantera data själva, utan vi behöver hjälp.



**Vad annat än temperaturen
kan påverka
glassförsäljningen?**



Vad mer påverkar glassförsäljningen?

Helg eller vardag

Regn eller solsken

Skollov

Evenemang i staden

Pris och kampanjer

Konkurrenter i närheten

Glassutbud

Sociala medier

Vad mer påverkar glassförsäljningen?

Helg eller vardag

Regn eller solsken

Skollov

Evenemang i staden

Pris och kampanjer

Konkurrenter i närheten

Glassutbud

Sociala medier

1 variabel: en linje.

2 variabler: ett plan.

T.ex. 200 variabler: omöjligt för en människa.

Vad mer påverkar glassförsäljningen?

Helg eller vardag

Regn eller solsken

Skollov

Evenemang i staden

Pris och kampanjer

Konkurrenter i närheten

Glassutbud

Sociala medier

1 variabel: en linje.

2 variabler: ett plan.

T.ex. 200 variabler: omöjligt för en människa.

*Det är där maskininlärning kommer in
– datorn hittar mönstret åt oss.*

Från glasskiosken till världen

På samma sätt ligger matematik bakom AI-system inom till exempel...



Spotify

rekommenderar låtar baserat på dina lyssnarvanor.



Google Translate

lär sig översätta från miljarder texter



Medicinsk diagnostik

hittar tumörer på röntgenbilder snabbare än läkare.



Sociala medier

väljer vilka inlägg du får se i ditt flöde.



Självkörande bilar

förutspår hur andra bilar kommer att röra sig.



Väderprognoser

kombinerar tusentals datakällor till en prognos.

Det viktigaste att ta med sig...

1

Modeller är gissningar med struktur

En modell tar in information och gör en förutsägelse. Den är aldrig perfekt.

2

Mer data = bättre modell

Med temperatur kan vi gissa mycket bättre än med bara medelvärdet.

3

Verkligheten är brusig

Ingen modell kan passa alla punkter, utan det finns alltid en felmarginal.

4

AI "lärt sig" genom att justera

Precis som ni gjorde – mer data, ny kalibrering, mindre fel.

5

AI bygger på matematik

Bygger på att förutspå utifrån tidigare data.

