

LINEAR REGRESSION

(Matematik grundforløb)

JACOB DEBEL

Created: 2019-09-05 Thu 11:38

PUNKTER OG RETTE LINJER

HVAD VED I FORVEJEN

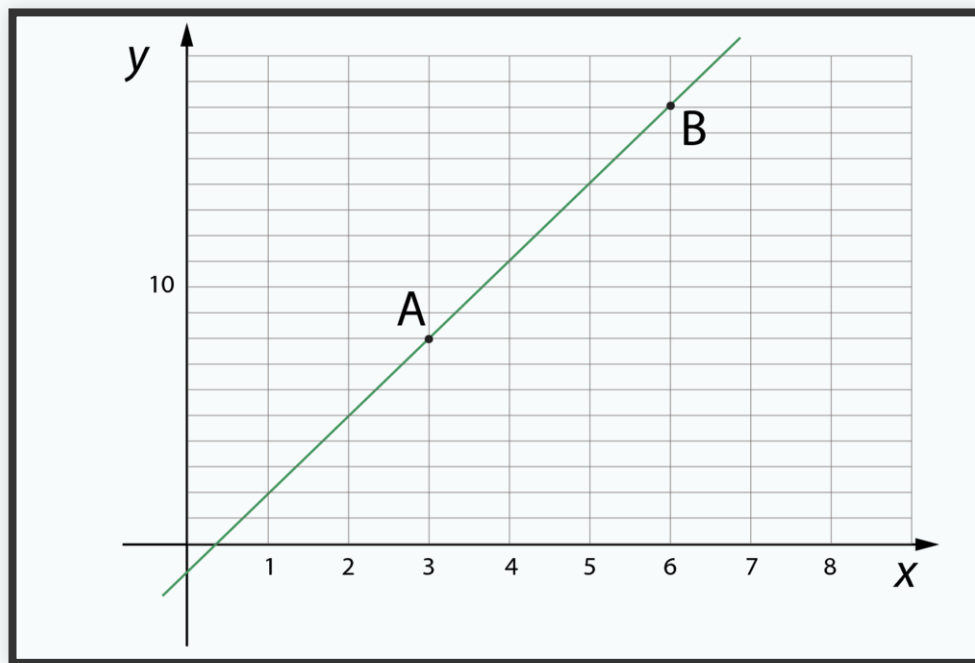
- Hvor mange punkter skal man kende, for at kunne tegne en ret linje?

HVAD VED I FORVEJEN

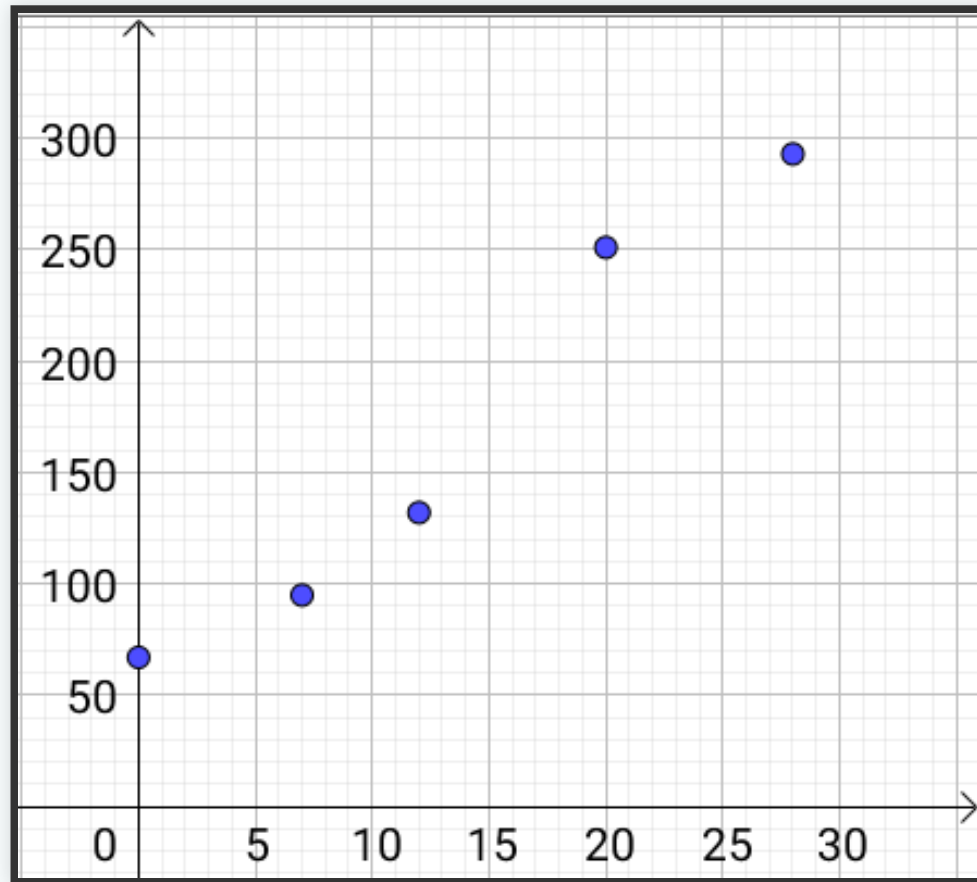
- Hvor mange punkter skal man kende, for at kunne tegne en ret linje?
 - 2

HVAD VED I FORVEJEN

- Hvor mange punkter skal man kende, for at kunne tegne en ret linje?
 - 2



HVAD HVIS DER ER FLERE END TO PUNKTER?



TEGN I HÅNDEN

Antal oddere i Danmark

Årstal, x	1984 (0)	1991 (7)	1996 (12)	2004 (20)	2012 (28)
Observationer, y	67	95	132	251	293

TEGN I HÅNDEN

Antal oddere i Danmark

Årstal, x	1984 (0)	1991 (7)	1996 (12)	2004 (20)	2012 (28)
Observationer, y	67	95	132	251	293

- Indtegn datasættet på tegnet papir i hånden.

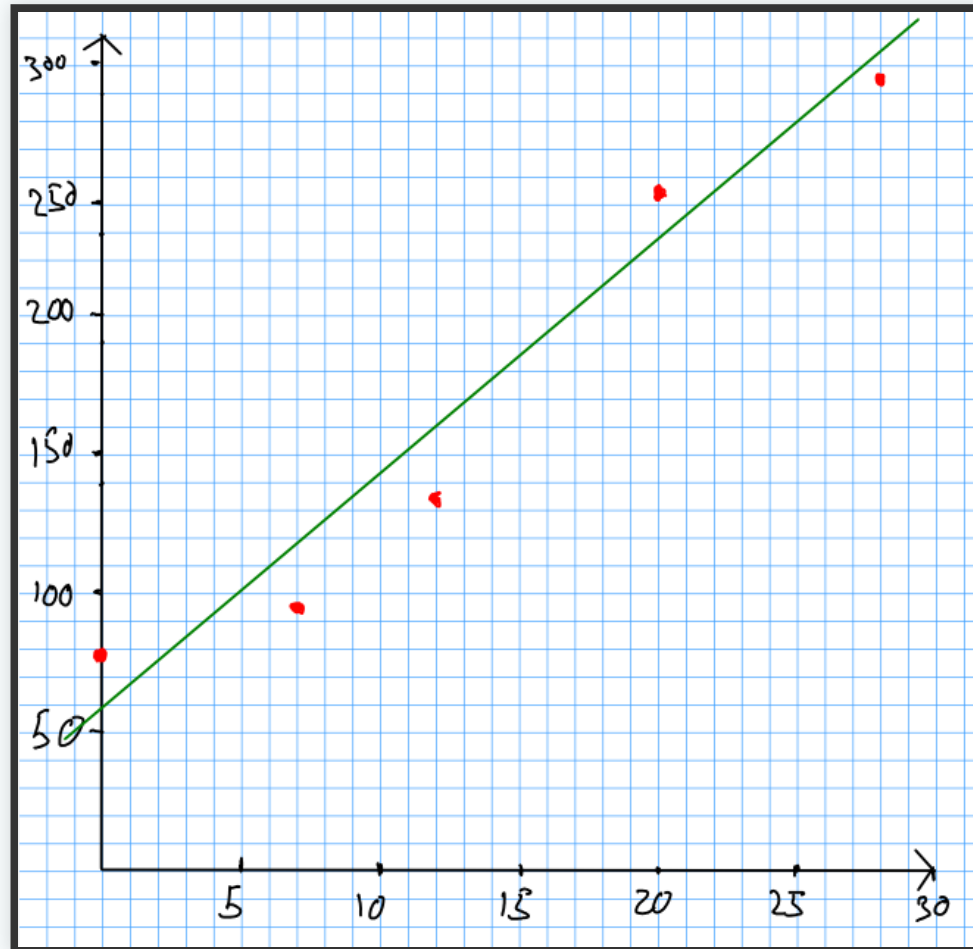
TEGN I HÅNDEN

Antal oddere i Danmark

Årstal, x	1984 (0)	1991 (7)	1996 (12)	2004 (20)	2012 (28)
Observationer, y	67	95	132	251	293

- Indtegn datasættet på tegnet papir i hånden.
- Indtegn selv den *bedst mulige* rette linje.

TEGN I HÅNDEN



KONKLUSION

KONKLUSION

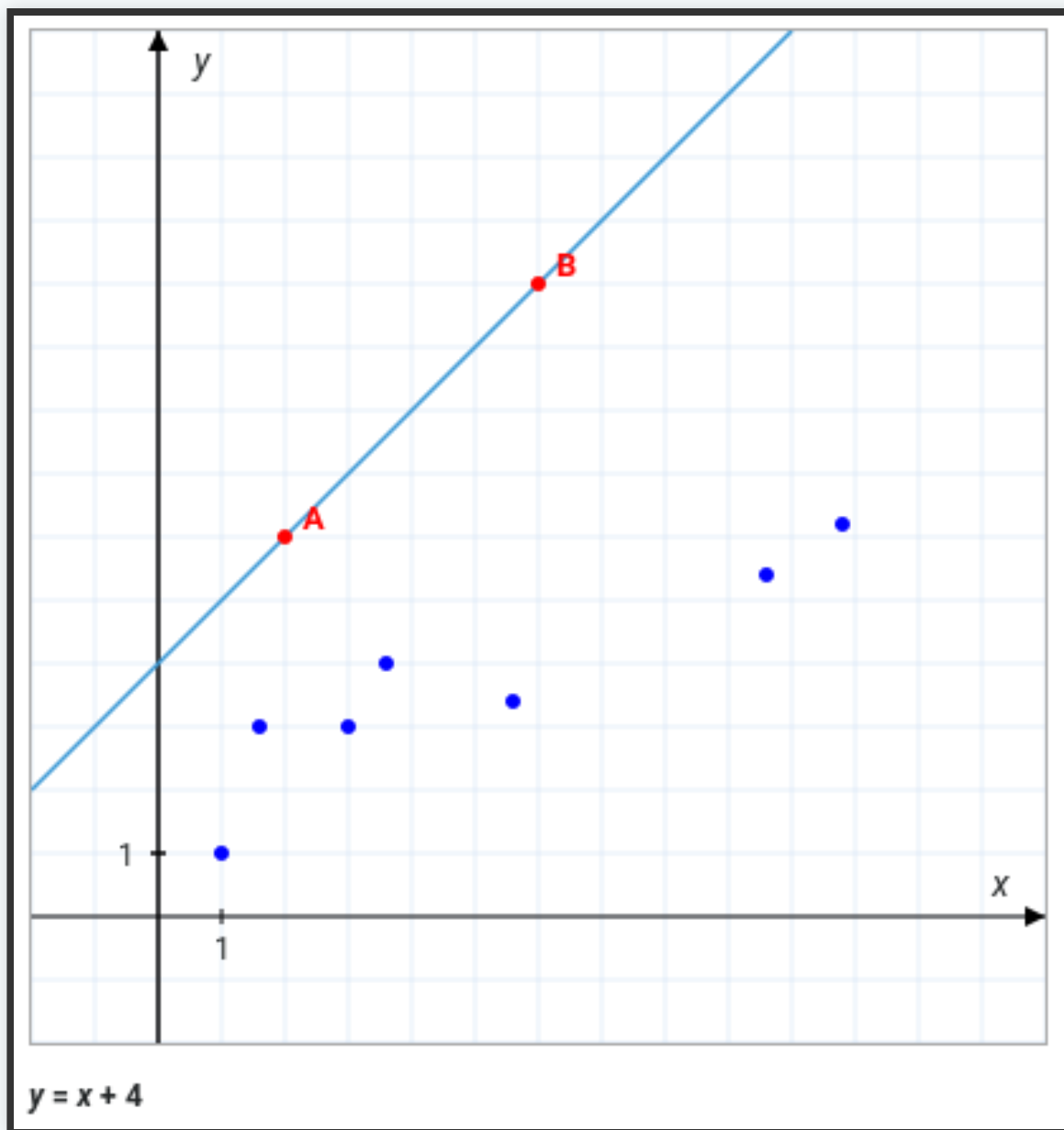
- Grafen går måske ikke igennem nogen af punkterne.

KONKLUSION

- Grafen går måske ikke igennem nogen af punkterne.
- Grafen skal være så tæt på alle punkterne som muligt.

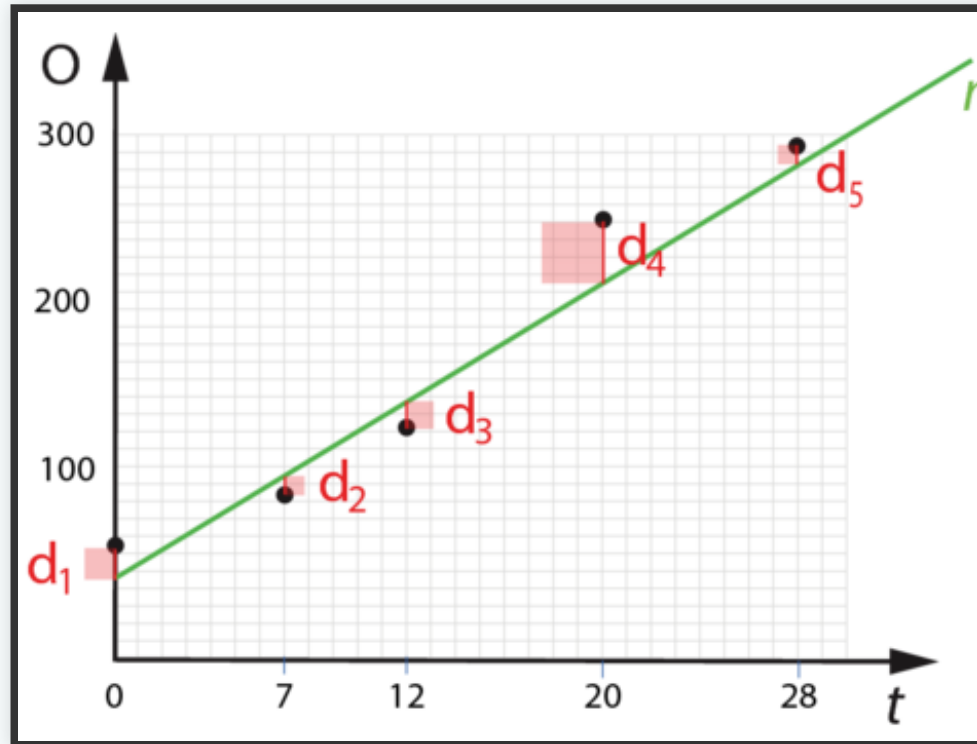
INTERAKTIV ØVELSE

<https://matbhtx.systime.dk/?id=c12359>

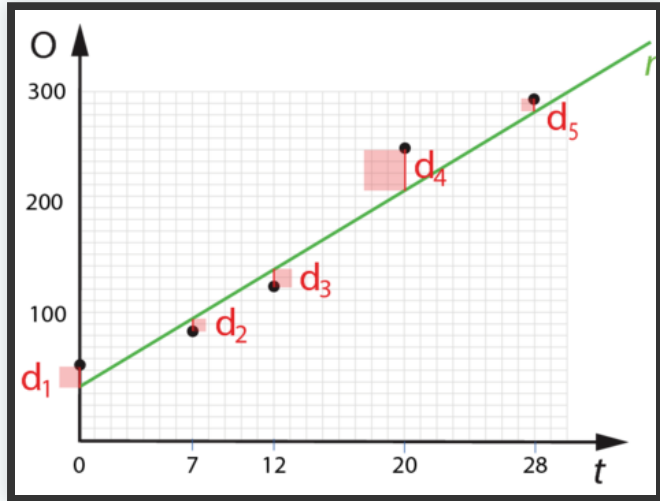


MINDSTE KVADRATERS METODE

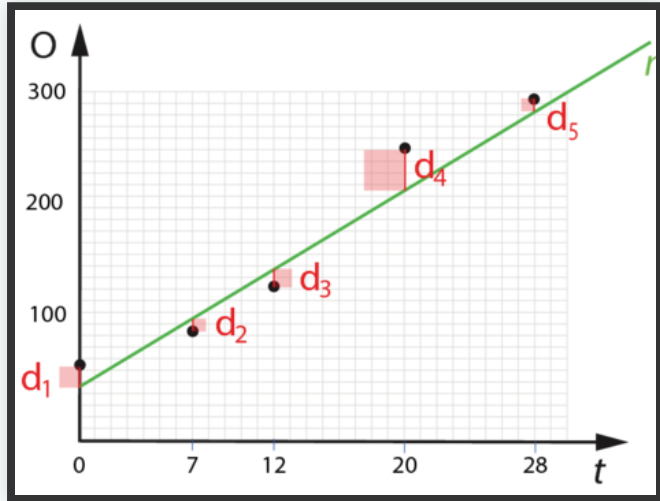
DEN BEDST MULIGE RETTE LINJE



DEN BEDST MULIGE RETTE LINJE

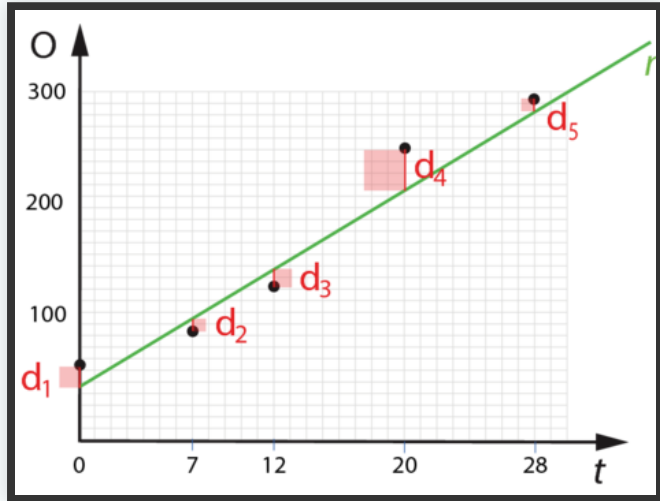


DEN BEDST MULIGE RETTE LINJE



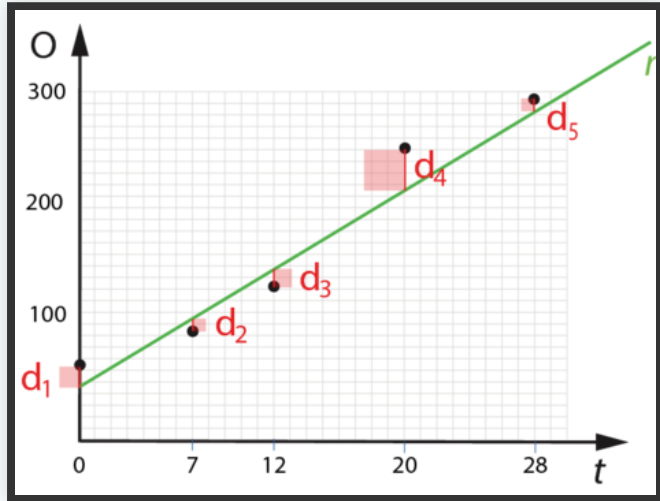
- Undersøger den lodrette afstand mellem de kendte punkter og en given ret linje.

DEN BEDST MULIGE RETTE LINJE



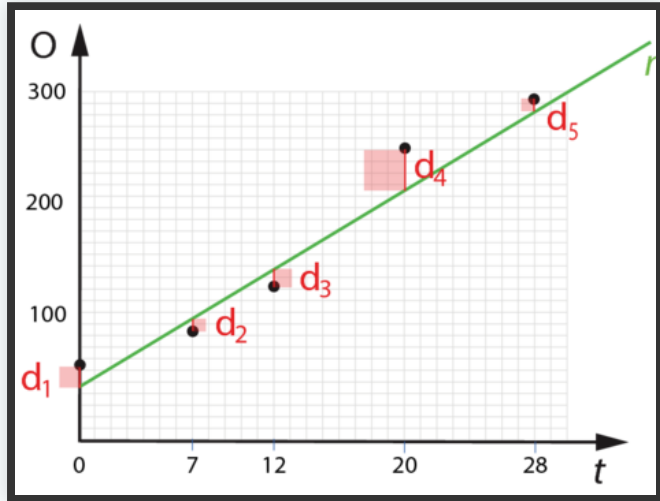
- Undersøger den lodrette afstand mellem de kendte punkter og en given ret linje.
- *Residualer*

DEN BEDST MULIGE RETTE LINJE



- Undersøger den lodrette afstand mellem de kendte punkter og en given ret linje.
- *Residualer*
- *Kvadratiske afvigelse*

DEN BEDST MULIGE RETTE LINJE

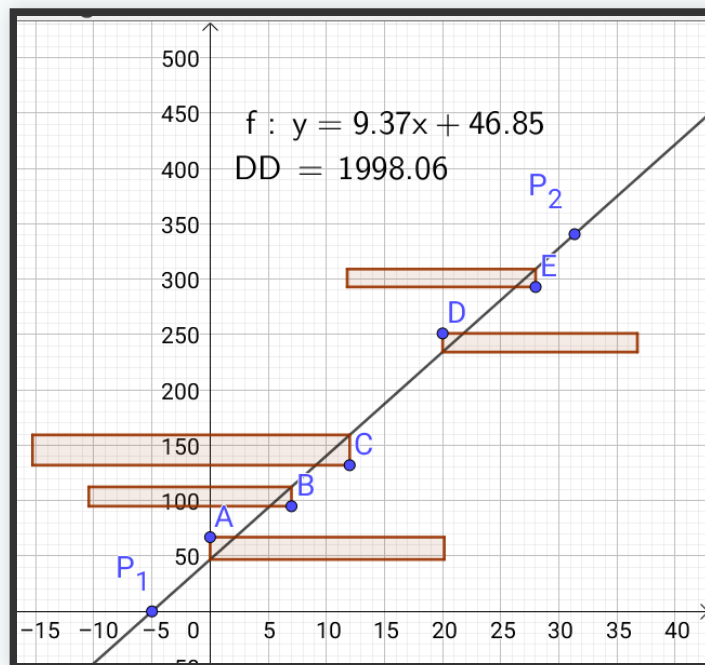


- Undersøger den lodrette afstand mellem de kendte punkter og en given ret linje.
- *Residualer*
- *Kvadratiske afvigelser*
- Summen af alle de kvadratiske afvigelser, skal være så lille som muligt.

$$D = d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

PRØV DET SELV

Mindste_kvadraters_metode.ggb på lectio



- Flyt på P_1 og P_2 .
- Få DD til at blive så lille som muligt.

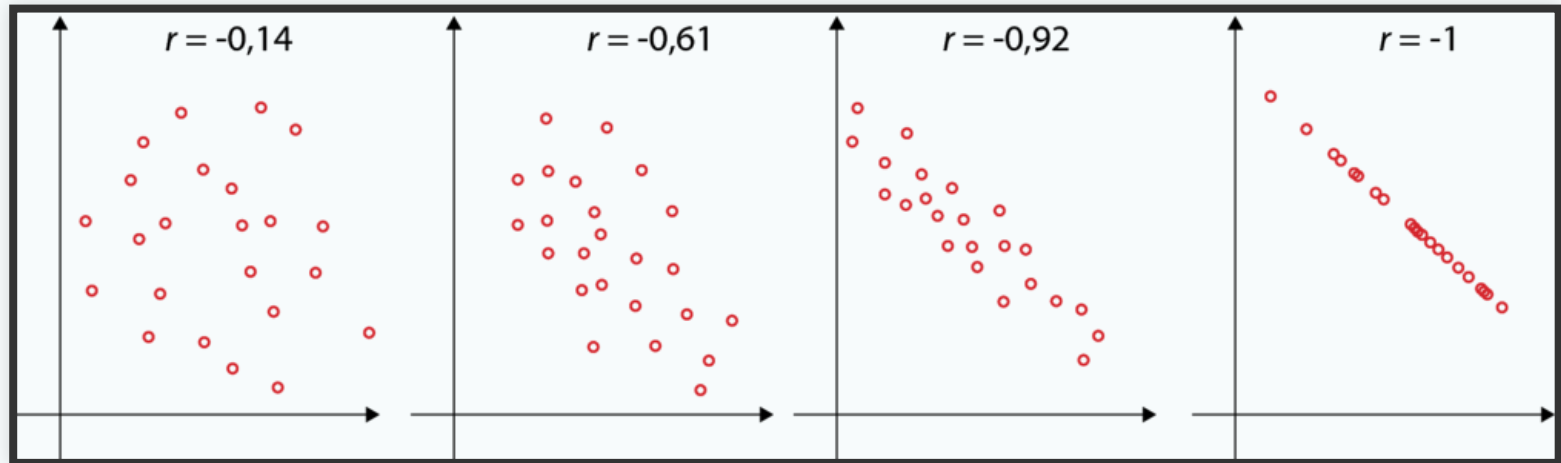
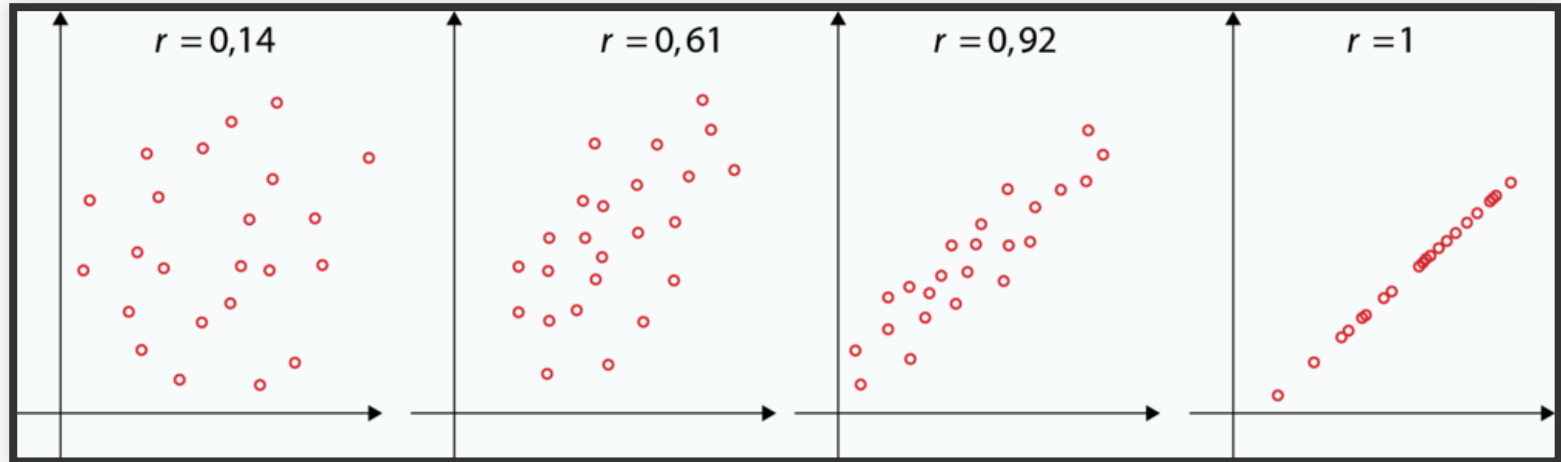
FÅ GEOGEBRA TIL AT GØRE DET

Indtast følgende i *input*-feltet:

- `liste1={A,B,C,D,E}` (Opret en liste med punkter)
- `FitLinje(liste1)` (Finder den bedste rette linje)

HVOR GODT PASSER DET
SÅ?

KORRELATIONSKOEFFICIENTEN, r



- Kommando i geogebra: `PMCC(<liste med punkter>)` (say what!?)

FORKLARINGSGRADEN, r^2

- $0 \leq r^2 \leq 1$
- $r^2 = 1$ - Perfekt *fit*.
- $r^2 = 0$ - Det dårligste fit overhovedet.
- $r^2 > 0.99$ - Godt lineært fit.
- $0.95 < r^2 < 0.99$ - Rimeligt lineært fit.
- Kommando i geogebra:

`Rkvadreret(<liste med punkter>,
<funktion>)`

OPGAVE

 OPGAVE 

Bestem med værktøjet den bedste rette linje for følgende talværdier.

x	2	4	6	8	10
y	3,3	3,7	3,9	4,1	4,8

- Indsæt punkter i geogebra
($(2, 3.3)$ i inputfeltet)
- Opret liste med punkterne
(`liste1={A,B}` og alle
andre punkter også)
- Fit en ret linje
(`fitlinje(liste1)`)
- Find forklaringsgraden
(`Rkvadreret(liste1)`)

FLERE OPGAVER

<https://matbhtx.systime.dk/index.php?id=1295>



OPGAVE 0.38 ID

Bestem regressionslinjen for følgende talpar.

x	1	2	3	4	5	6
y	8	9	11	11	12	13



OPGAVE 0.39 ID

Bestem regressionslinjen for følgende talpar.

x	10	20	30	40	50	60
y	8	9	11	11	12	13



OPGAVE 0.40 ID

Bestem regressionslinjen for følgende talpar.

x	-1	7	9	22	77	100
y	-4	2	11	33	81	113



OPGAVE 0.41 ID

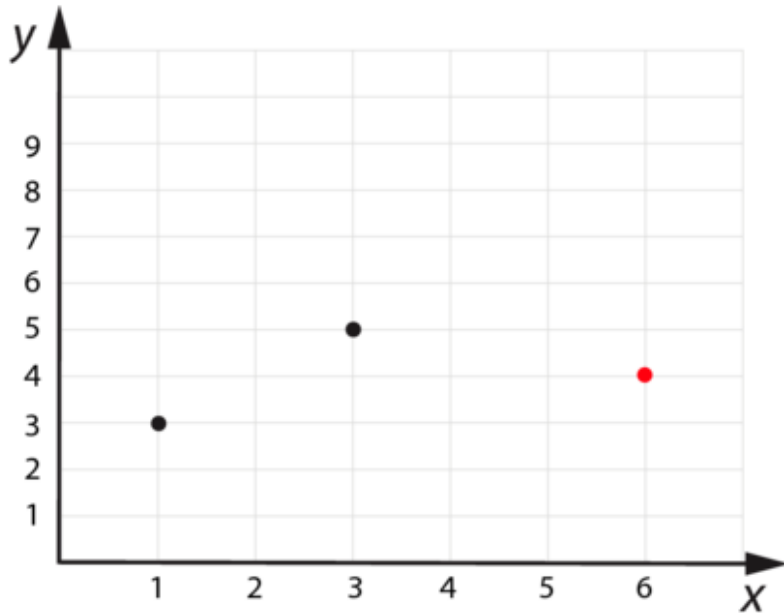
Bestem regressionslinjen for følgende talpar.

x	1	2	3	4	5	6
y	-1	-2	-11	-5	-17	-22

FLERE OPGAVER



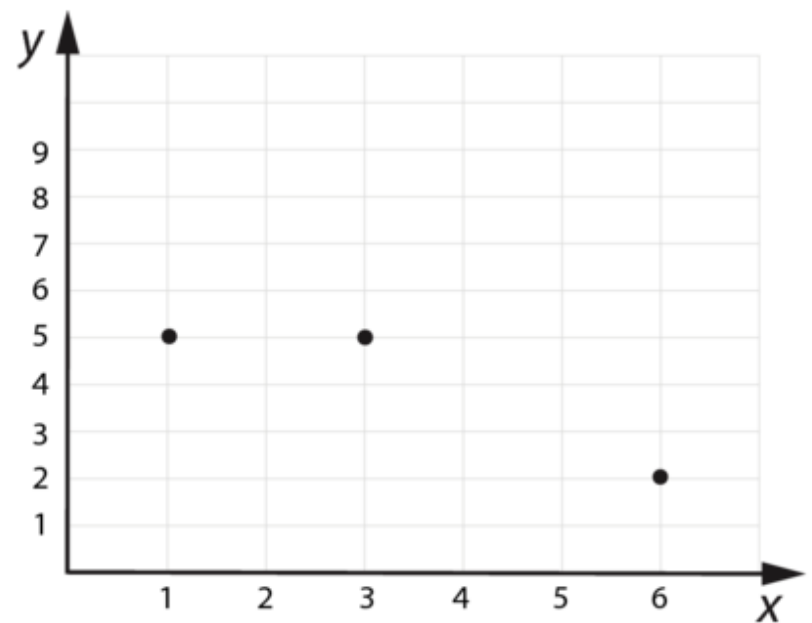
OPGAVE 0.43



Hvilken y -værdi skal det røde punkt $(6, y)$ have, for at korrelationsfaktoren for de tre punkter bliver lig med 1?



OPGAVE 0.44



Bestem regressionslinje og korrelationsfaktor for punkterne i koordinatsystemet.

BACKUP SLIDES

HVIS MAN SKAL REGNE DET HELE I HÅNDEN

2.2.1 Bestemmelse af forskrift

Den bedst mulige rette linje, \hat{y} , bestemmes på følgende måde:

$$\hat{y} = \hat{a} \cdot x + \hat{b}. \quad (2.1)$$

\hat{a} og \hat{b} er henholdsvis det estimerede hældningstal og skæring med y-aksen.

\hat{a} og \hat{b} bestemmes via følgende formler:

$$\hat{a} = \frac{\sum_{i=1}^n (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2.2)$$

n er antallet af datapunkter, \bar{x} er gennemsnittet af de målte x-værdier og \bar{y} er gennemsnittet af de målte y-værdier.

$$\hat{b} = \bar{y} - \hat{a} \cdot \bar{x}, \quad (2.3)$$

HVIS MAN SKAL REGNE DET HELE I HÅNDEN

2.2.2 Forklaringsgrad

Til at vurdere, hvor godt den estimerede rette linje passer med datapunkterne, benyttes *korrelationskoefficienten*, r , og den nært beslægtede *forklaringsgrad*, r^2 . r^2 bestemmes på følgende måde:

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.4)$$

Hvis $r^2 = 1$ går den bedste rette linje igennem *alle* punkterne. Hvis $r^2 = 0$ er der absolut ingen tendens til, at punkterne danner en ret linje.