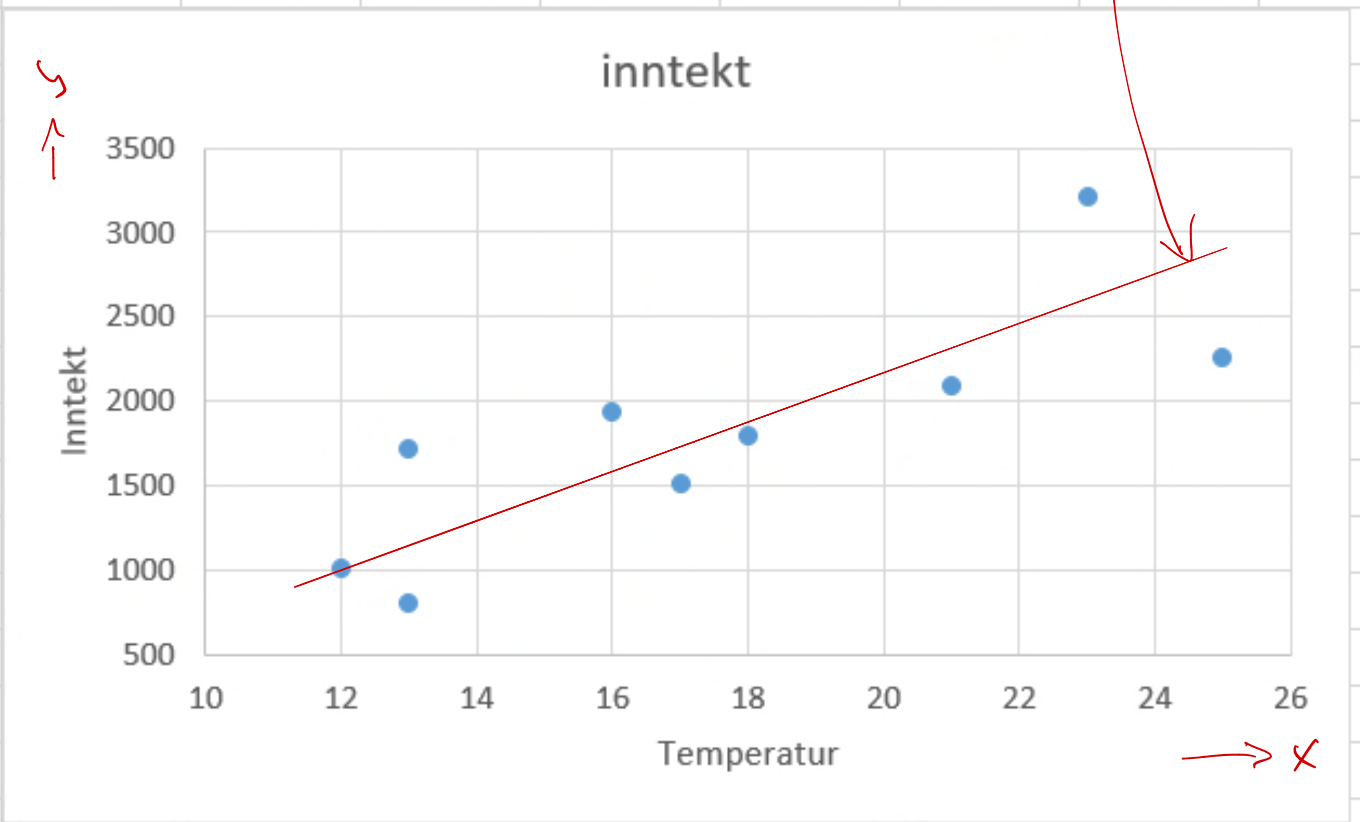
# LINEÆR REGRESJON: EKSEMPEL

Anta vi ser på inntekten til en is-kiosk pr. dag.
Anta vi har følgende data

| Dag | Temp ($x$) | inntekt ($y$) |
|---|---|---|
| 03.jun | 23 | 3213 |
| 10.jun | 21 | 2089 |
| 17.jun | 25 | 2253 |
| 24.jun | 18 | 1801 |
| 01.jul | 13 | 801 |
| 08.jul | 16 | 1934 |
| 15.jul | 13 | 1720 |
| 22.jul | 17 | 1514 |
| 29.jul | 12 | 1017 |

Linear sammenheng: VI ANTAR linear samm.

$$\hat{y} = \hat{a} + \hat{b}x$$



# LINEÆR REGRESJON: EKSEMPEL

$\hat{b}$ = stigningstall $= \frac{S_{xy}}{S_x^2} = \frac{2659.64}{21.5278} = 123.545$

$$\hat{y} = \overbrace{-353.115}^{\hat{a}} + \overbrace{123.545}^{\hat{b}}\, x$$

$\hat{a}$ = skj.pkt med y-aksen

$= \bar{Y} - \hat{b}\,\bar{X} = 1815.78 - 123.545 \cdot 17.556$

$= -353.115$

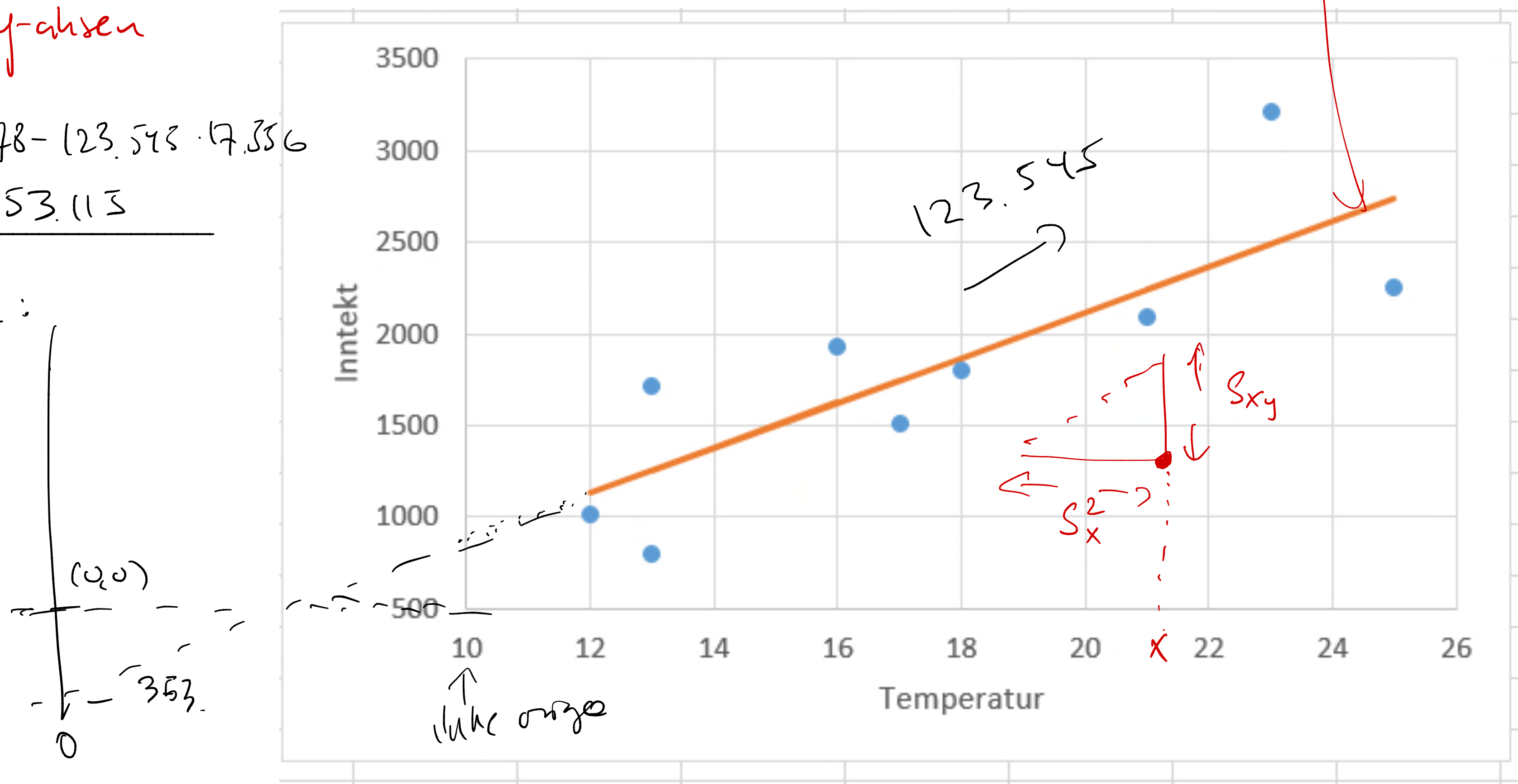
Oppgitt som regel:

$S_x^2 = 21.5278$

$S_{xy} = 2659.64$

$\bar{Y} = 1815.78$

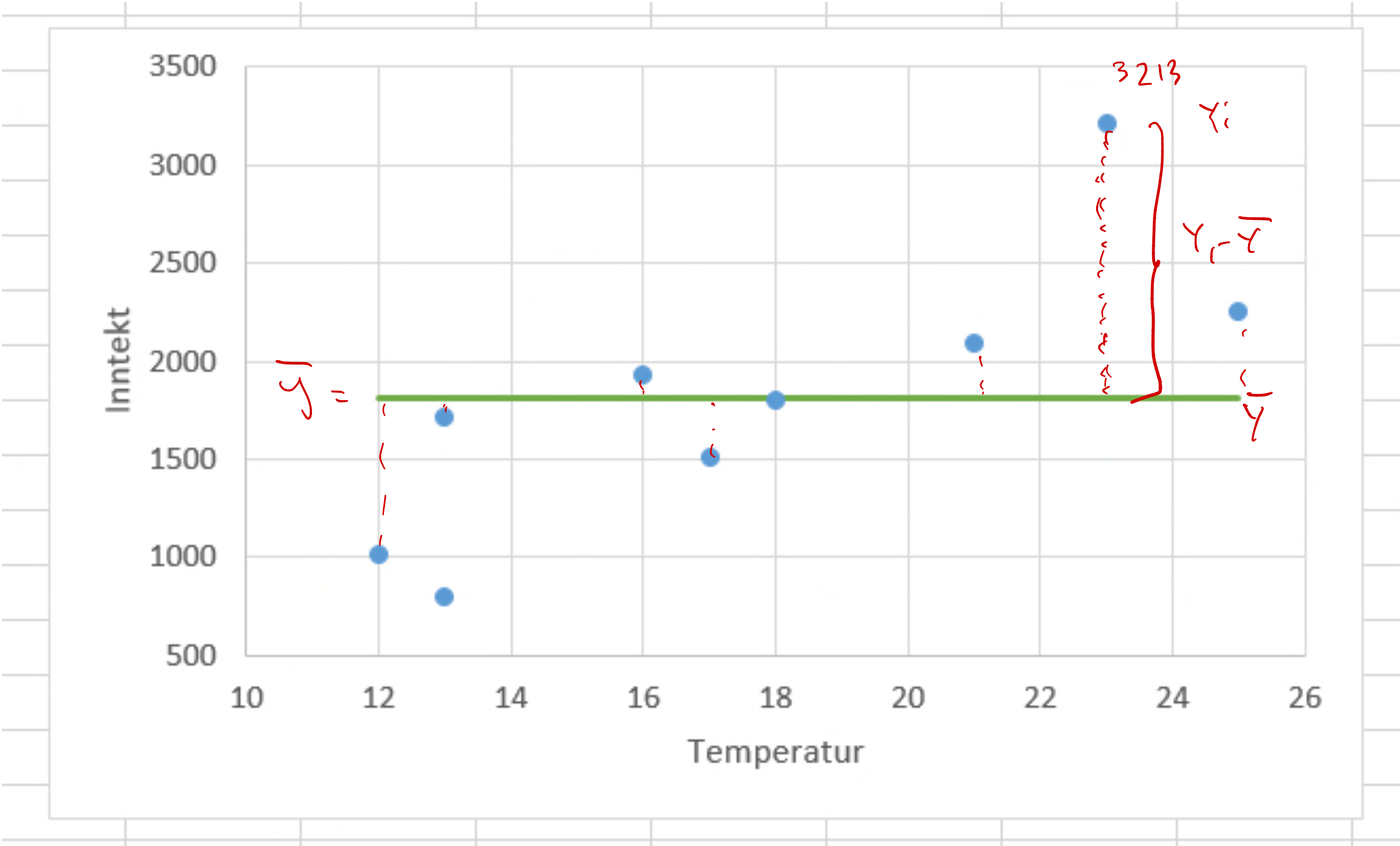
$\bar{X} = 17.556$



# FORKLARINGSKRAFT

Vi har : $\bar{Y} = 1815.78$

Vi prøver å FORKLARE VARIANSEN rundt $\bar{Y}$.



Vi ønsker med REGRESJON å forklare noe av variansen rundt $\bar{Y}$.

$$\hat{y} = -353.115 + 123.545x$$

Fra figuren:

$$Y_i - \bar{Y} = \underbrace{(y_i - \hat{y}_i)}_{\text{ikke forklart varians}} + \underbrace{(\hat{y}_i - \bar{y})}_{\text{Forklart varians}}$$

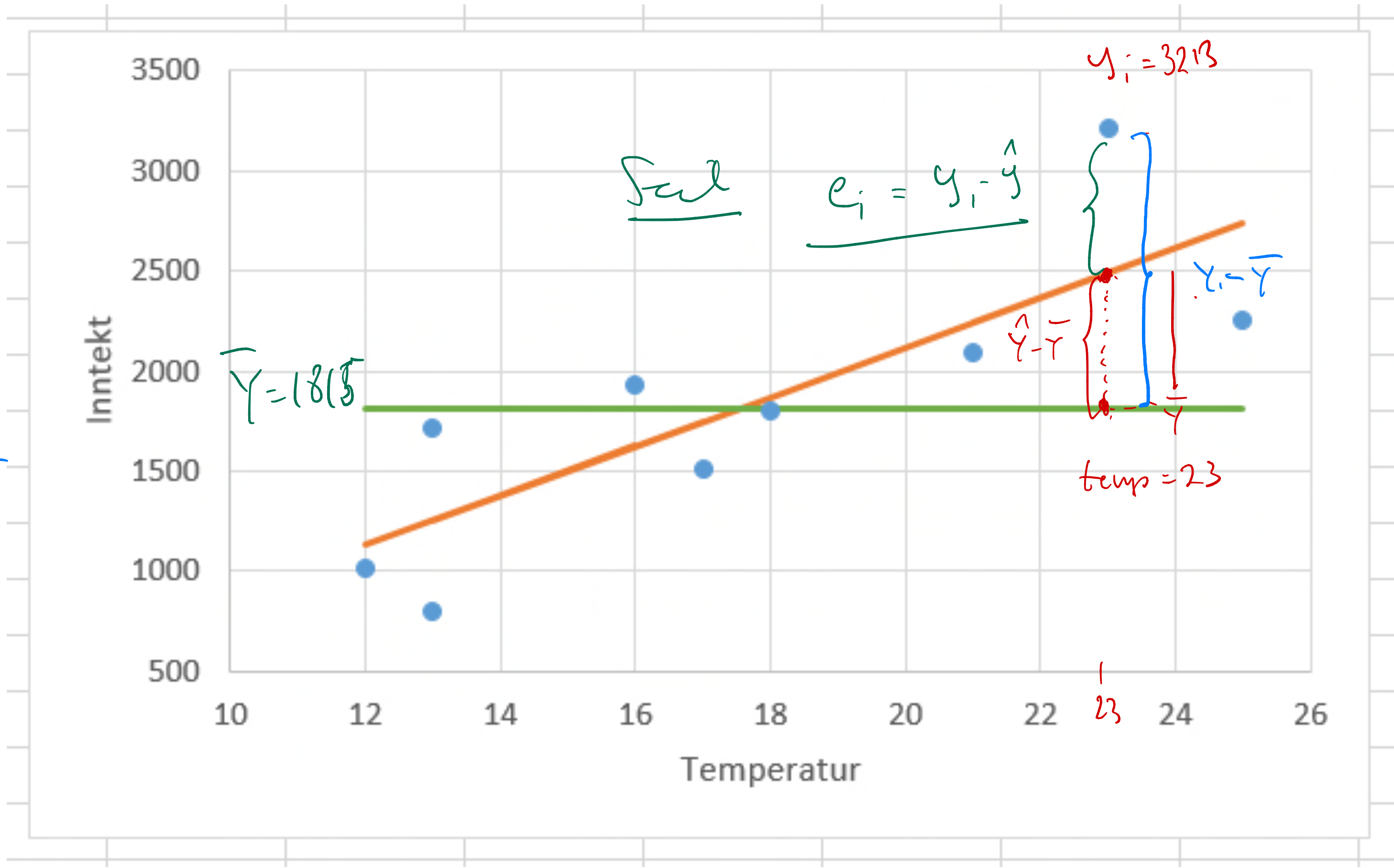
$$\sum_{i=1}^{n=9} (y_i - \bar{y})^2 = \sum_{i=1}^{n=9} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n=9} (\hat{y}_i - \bar{y})^2$$

$$\text{SST} = \text{SSE} + \text{SSR}$$

SST = Avvik ift $\bar{Y}$

SSE = Feil ift $Y, \hat{Y}$

SSR = forklart varians



Konklusjon: (VARIANSANALYSE)

$$\frac{\overbrace{SST}^{1}}{SST} = \frac{SSE}{SST} + \frac{\overbrace{SSR}^{R^2}}{SST} \Rightarrow 1 = \frac{SSE}{SST} + R^2$$

$$\boxed{R^2 = 1 - \frac{SSE}{SST}}$$

$$4000301 = 1371631 + 2628670$$

Hvor stor ANDEL % av SST har vi forklart ved Regresjon (SSR).

$$\boxed{r^2 = R^2} = \text{FORKLARINGSKRAFT} = \frac{SSR}{SST} = \frac{\text{Forklart varians}}{\text{Total varians}} = \frac{2628670}{4000301}$$

$$= 0.657$$

komp.

Konklusjon:

$$R^2 = \frac{SSR}{SST} = \overbrace{1 - \frac{SSE}{SST}}^{\text{står i komp.}} = \boxed{0{,}657}$$

Typisk spm: Finn $R^2$

Typisk er SSE og SST oppgitt.

Neste spm: Tolk $R^2 = 0{,}657$ $\xrightarrow{100\%}$ 65,7%

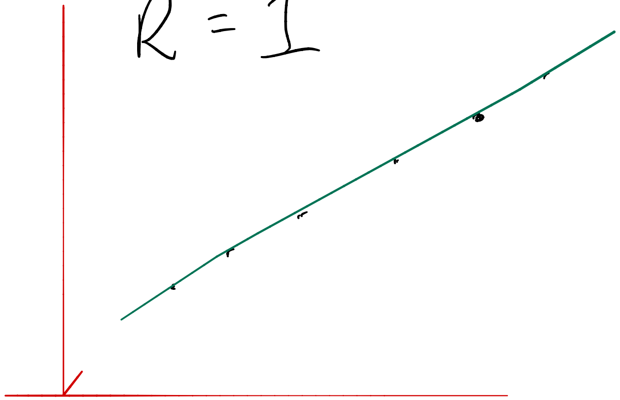
Regresjonslinjen $\hat{y} = -353 + 123x$ forklarer 65,7% av den totale variansen. []

$$R^2 = \frac{SSR}{SST}$$

($R^2$ alltid mellom 0 og 1)

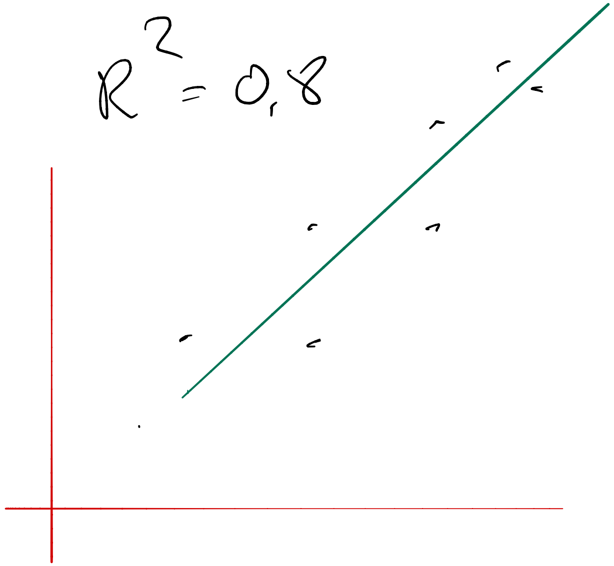
100% forklart

$R^2 = 1$

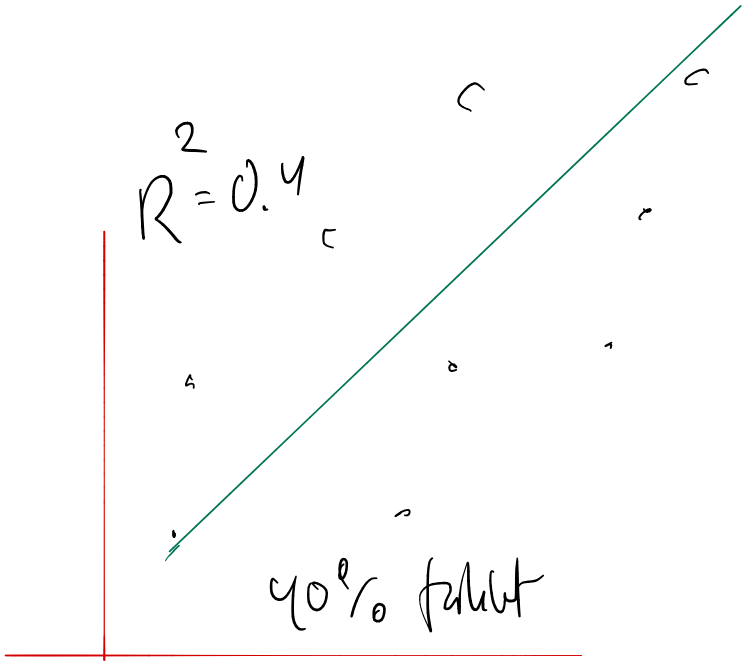


$R^2 = 0,8$

80% forklart

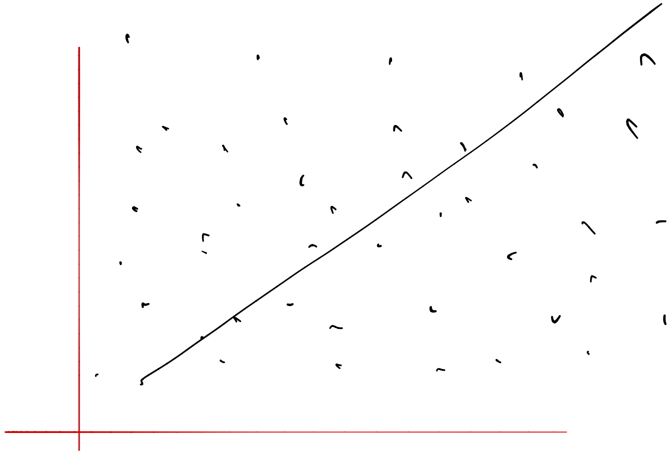


$R^2 = 0.4$



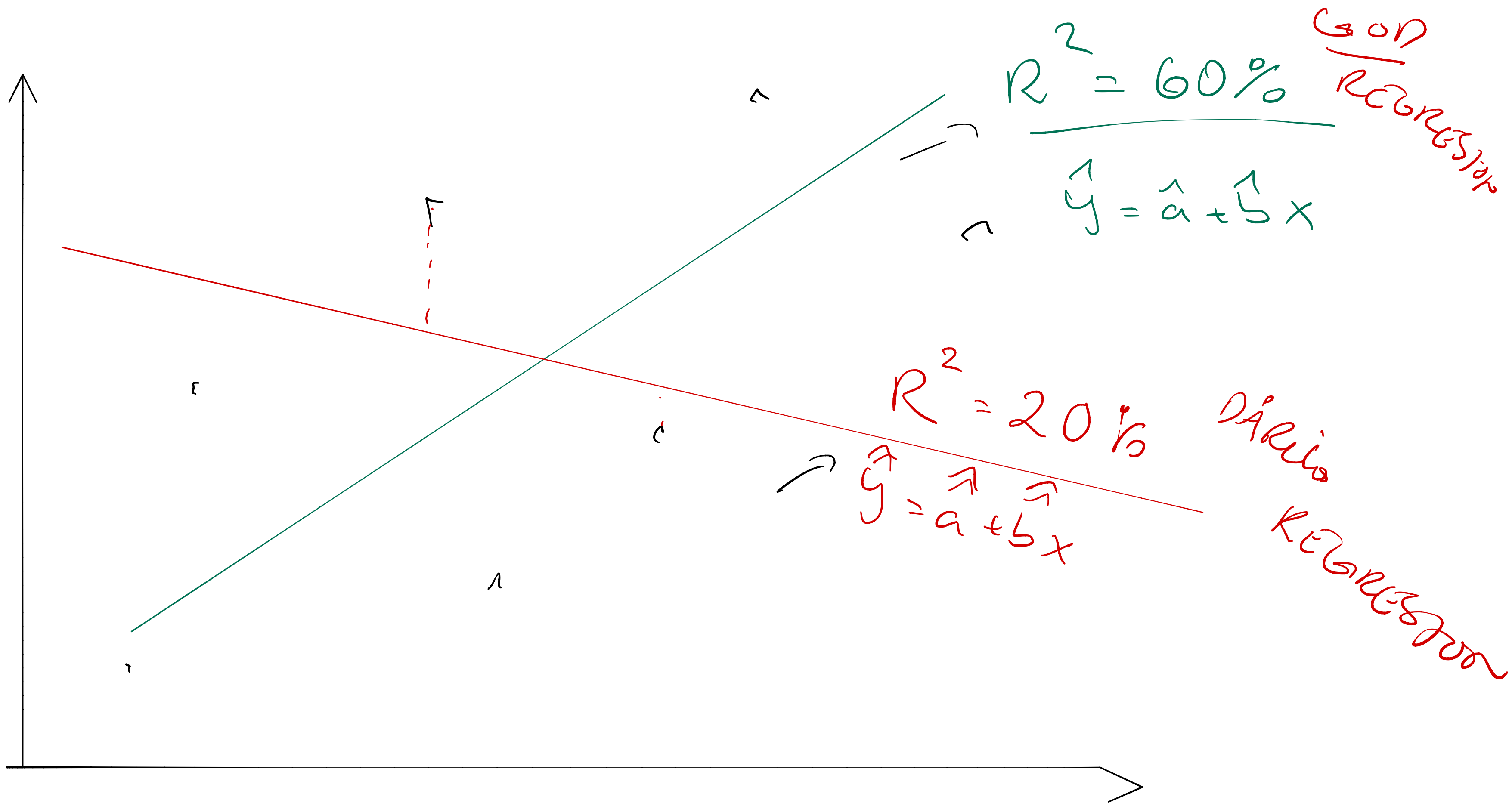
40% forklart

$R^2 \approx 0$



$R^2$ måler i hvilken grad vi har lineær sammenheng

$\underline{R^2}$ er koblet mot VÅR linje $\boxed{\hat{y} = \hat{a} + \hat{b}x}$




$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$ = korrelasjonskoeffisient (Sier bare __OM__ dataene har en lineær sammenheng, ikke hvilken linje)

$$0 \le R^2 \le 1$$ (En prosent)

$$\underbrace{-1}_{\text{nedadgående linje}} \le r_{xy} \le \underbrace{1}_{\text{oppadstigende linje}}$$ (en tendens)

## Logisk gang:

1) Data $(x_1, y_1) \dots (x_n, y_n)$

2) Finn $r_{xy} = \frac{S_{xy}}{S_x S_y}$ korrelasjonskoeff.

Hvis

$r_{xy} \approx 1$ eller $r_{xy} \approx -1$

så finnes en linje i datasettet.

3) Finn $\hat{y} = \hat{a} + \hat{b}x$

hvor

$$\left.\begin{aligned} \hat{b} &= \frac{S_{xy}}{S_x^2} \\ \hat{a} &= \bar{Y} - \hat{b}\bar{x} \end{aligned}\right\} \text{ minimerer } SSE = \sum_{i=1}^{n} e_i^2$$

4) Finn forklaringskraften til $\hat{y} = \hat{a} + \hat{b}x$

$$r^2 = R^2 = \frac{SSR}{SST} = \left(1 - \frac{SSE}{SST}\right) = \text{prosent } \%$$

Fra eksempel:

$$R^2 = 65,7\ \%$$

Dette sier noe om hvordan TEMPERATUREN forklarer inntekten til IS-SALGET.
(Beste linja vi kan finne)

→ For å få bedre tilpasning, så må vi finne NYE forklaringsvariabler.

→ F.eks PRIS → $\hat{y} = \hat{a} + \hat{b}x + \hat{c}z \implies R^2 = 75\ \%$

pris

MULTIPLE REGRESJON.