

Oversiktsforelesning

Thea Bjørnland

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon

Nøkkeltbegreper:

Gjennomsnitt

Median

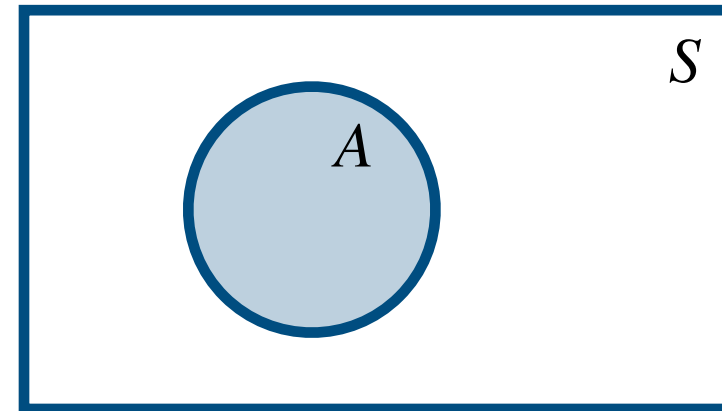
Standardavvik

Boksplott

Histogram

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon



Nøkkelbegreper:

Venndiagram

Snitt

Union

Komplementærhendelse

Sannsynlighet for
hendelser

Disjunkte hendelser

Betinget sannsynlighet

Uavhengige hendelser

Kombinatorikk

+ regneregler...

Uke 2: Hendelser og sannsynlighet

Eksempel: Terningkast

Utfallsrom: $S = \{1\text{-er}, 2\text{-er}, 3\text{-er}, 4\text{-er}, 5\text{-er}, 6\text{-er}\}$

Hendelser: $A = \{1\text{-er}\}$ $B = \{1\text{-er}, 2\text{-er}, 3\text{-er}\}$

$$P(A \cap B) = ?$$

Sannsynlighet for hendelser: $P(A) = ?$ $P(B) = ?$

$$P(A \mid B) = ?$$

Her: Alle enkeltutfall er like sannsynlige!

antall gunstige / antall mulige

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
- 3. Stokastiske variabler**
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon

Uke 3: Stokastiske variabler

Eksempel: Terningkast

$S = \{1\text{-er}, 2\text{-er}, 3\text{-er}, 4\text{-er}, 5\text{-er}, 6\text{-er}\}$

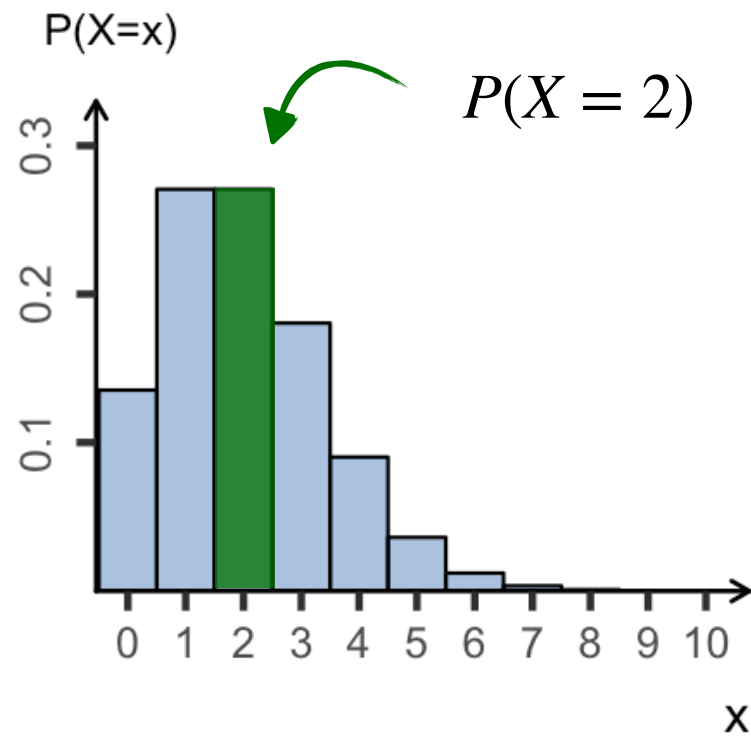
$X =$ antall øyne på terningen i ett kast

$$V_x = \{1, 2, 3, 4, 5, 6\}$$

Uke 3: Stokastiske variabler

Diskret

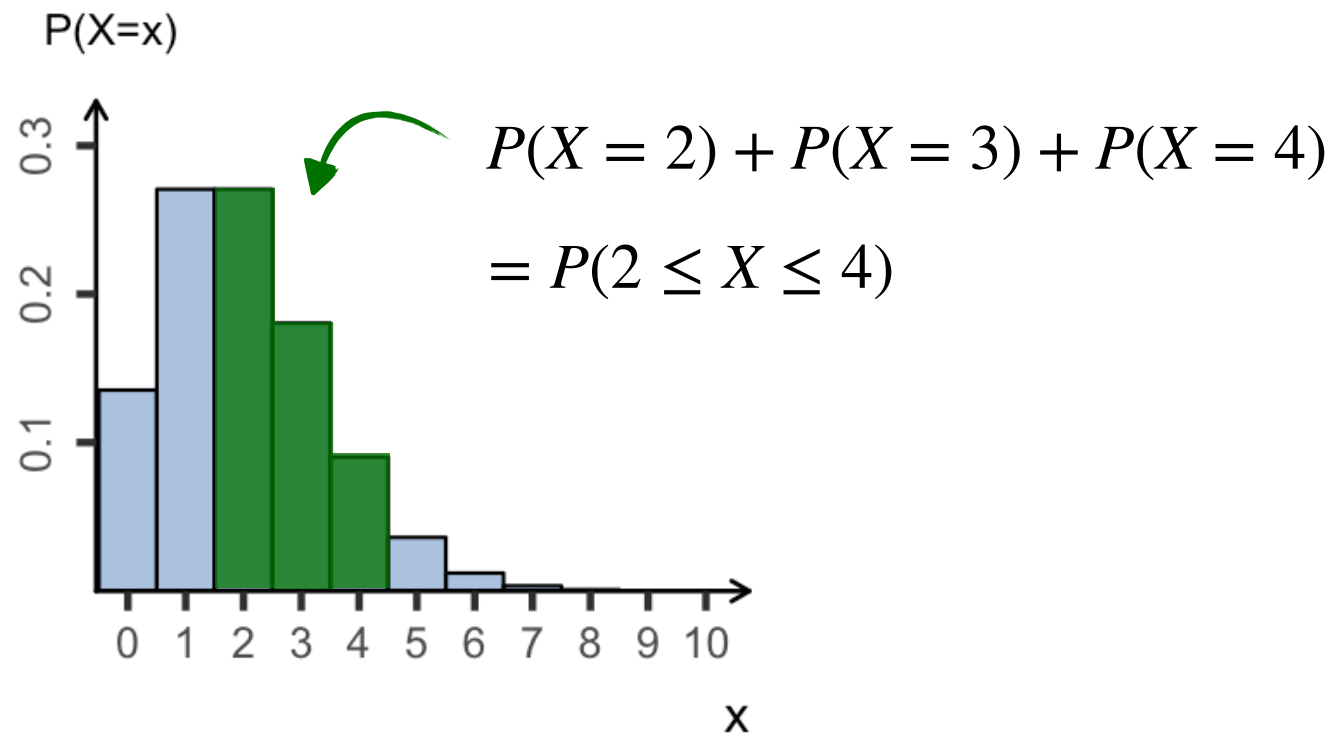
Kontinuerlig



Uke 3: Stokastiske variabler

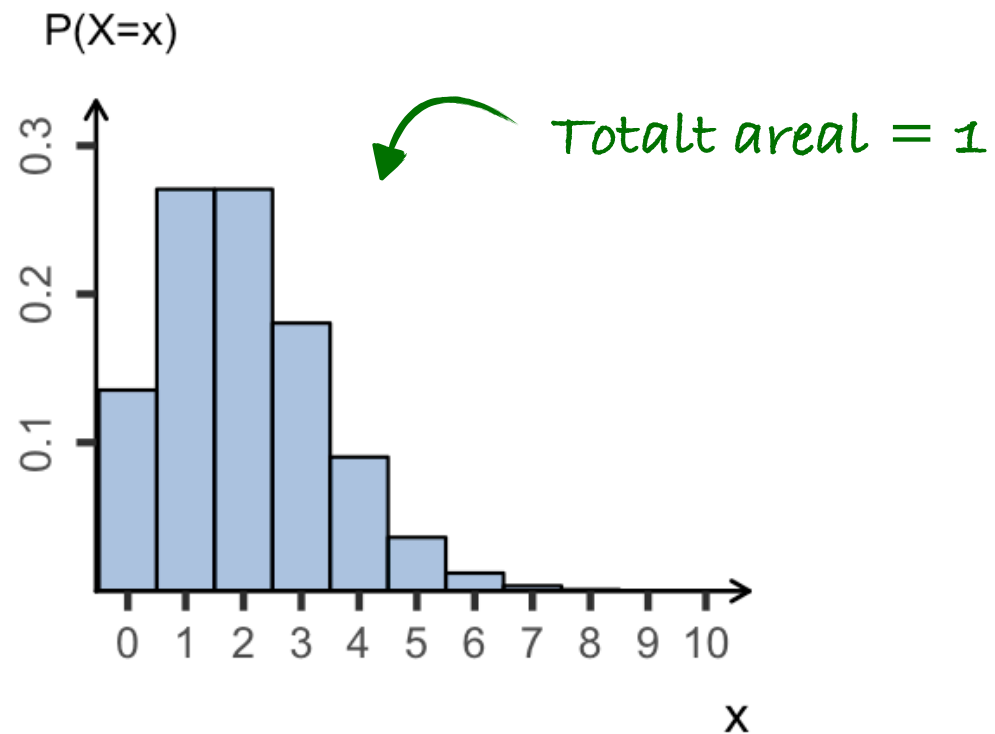
Diskret

Kontinuerlig

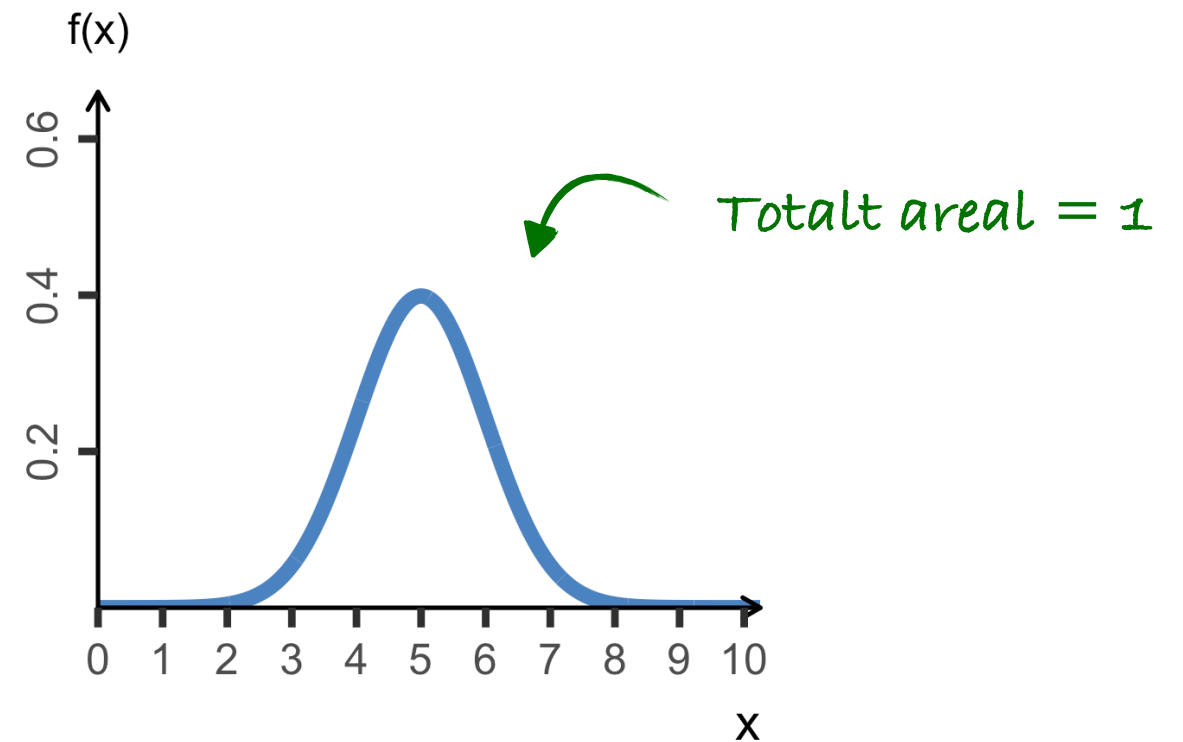


Uke 3: Stokastiske variabler

Diskret

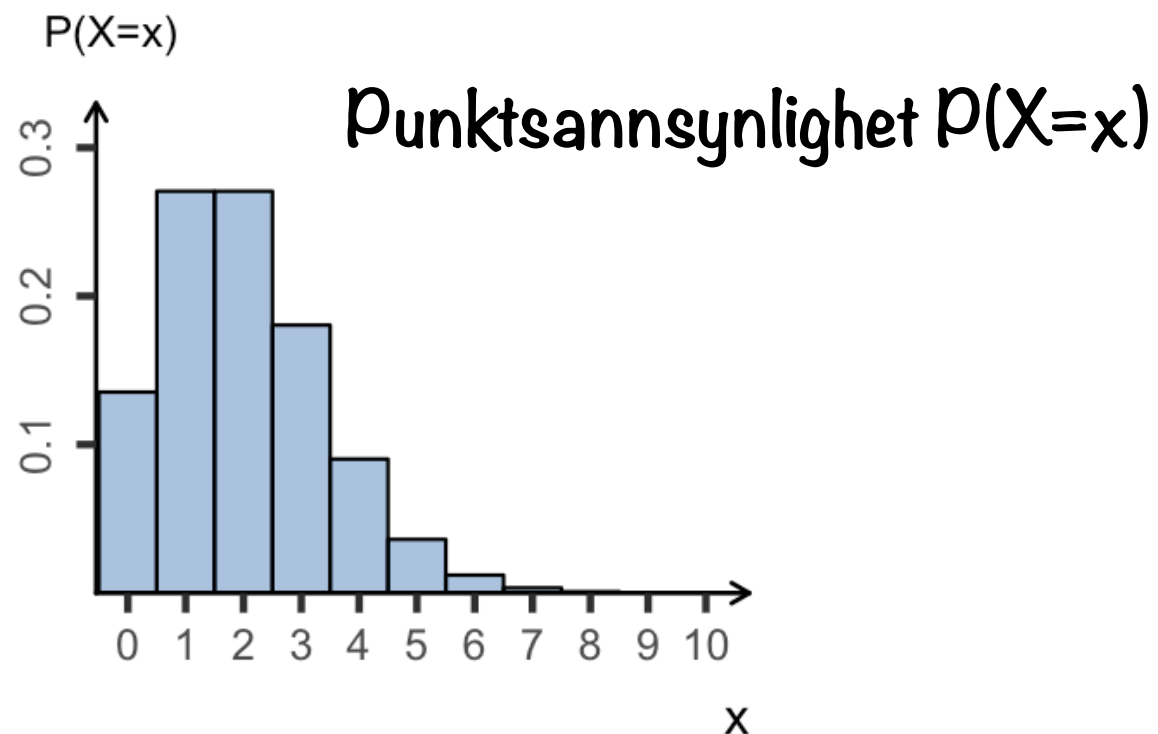


Kontinuerlig

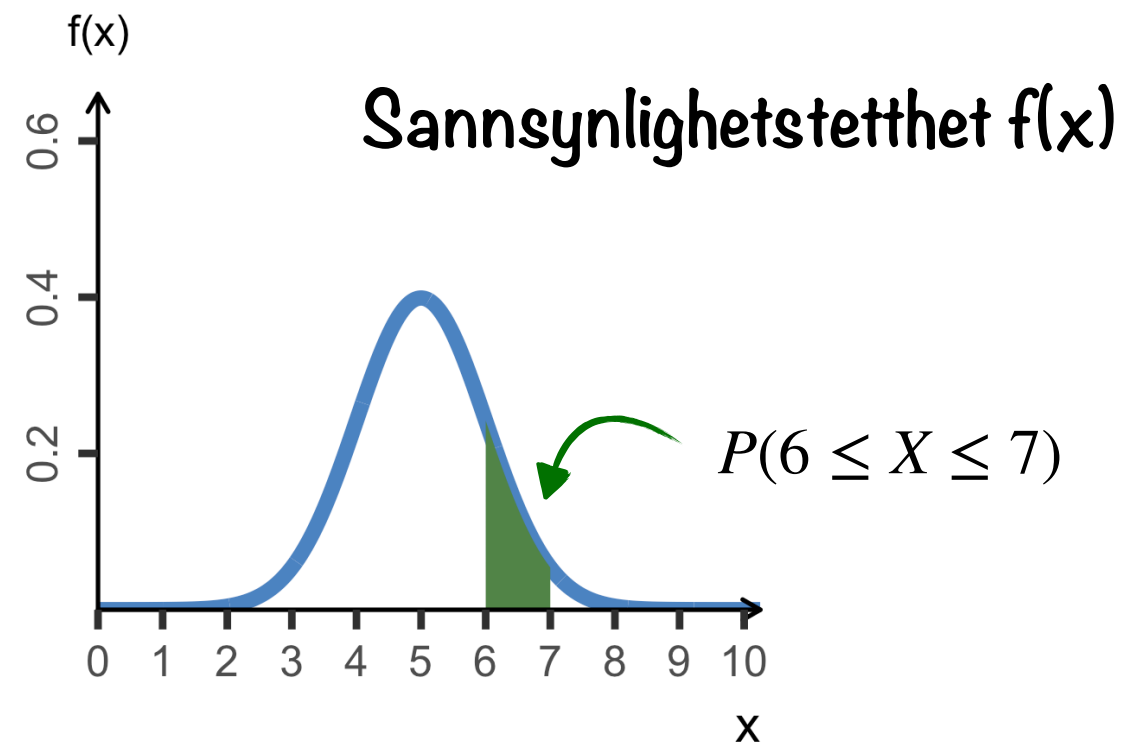


Uke 3: Stokastiske variabler

Diskret



Kontinuerlig



Nøkkelbegreper:

Kumulativ fordeling

$$F(x) = P(X \leq x)$$

Forventningsverdi $E(X)$

Kovarians

Varians $\text{Var}(X)$ og standardavvik $\text{SD}(X)$

Uavhengige stokastiske variabler

+ regneregler...

$$aX+b$$

$$X+Y$$

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon

Et forsøk kan ende som
suksess eller fiasko

Sannsynligheten for suksess, p ,
er den samme i alle forsøk

Forsøkene er uavhengige

X : antall suksesser i n forsøk

$$X \sim \text{Binomisk}(n, p)$$

Y : Antall forsøk frem til og med
første suksess

$$Y \sim \text{Geometrisk}(p)$$

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon

Hendelser skjer uavhengig
av hverandre

Rate: λ hendelser/tid

To hendelser kan ikke skje akkurat samtidig

X_t : antall hendelser i tidsintervall
av lengde t

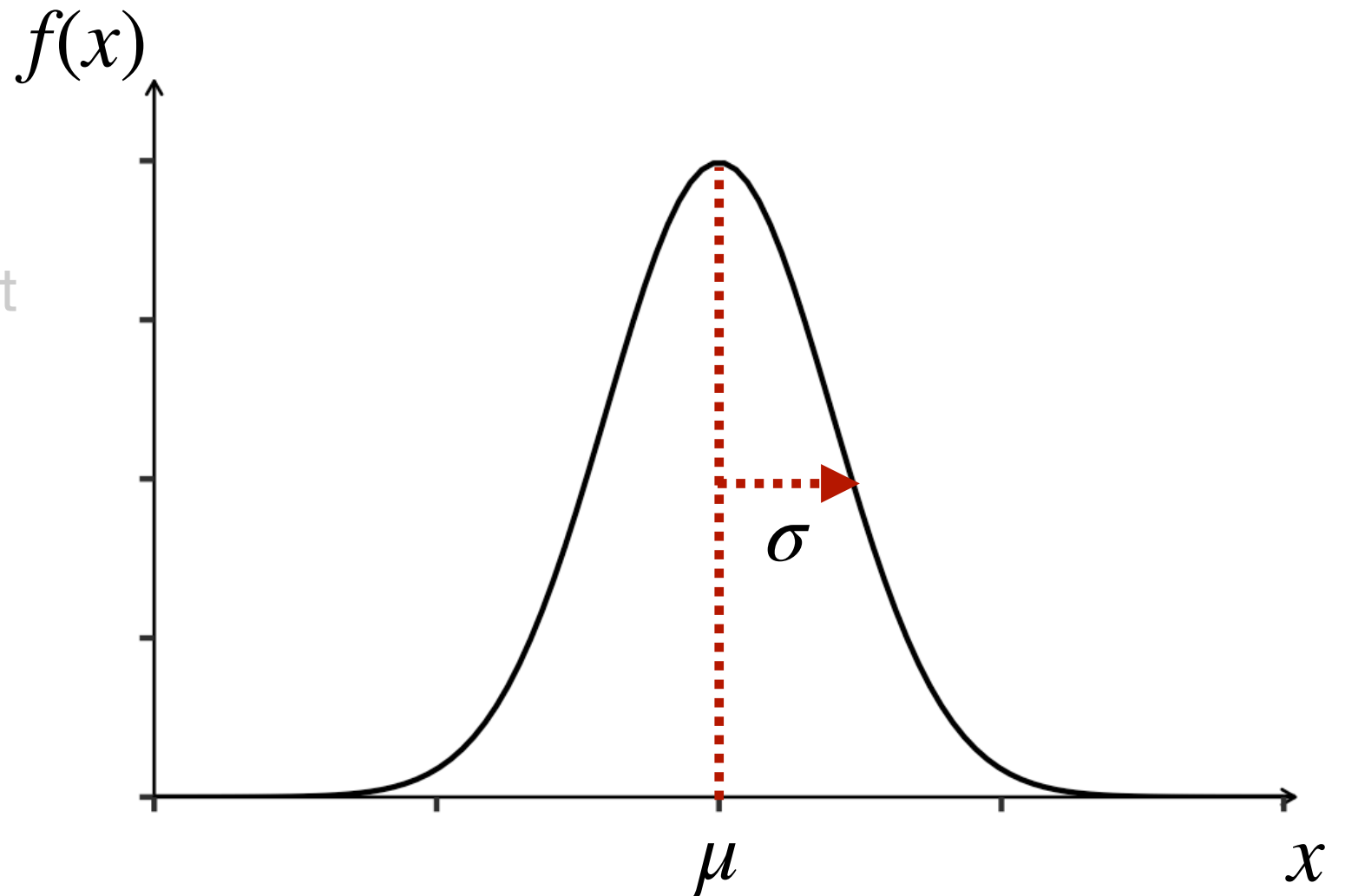
$$X_t \sim \text{Poisson}(\lambda t)$$

T : Tid frem til første/neste hendelse

$$T \sim \text{Eksponential}(\lambda)$$

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon



$X \sim N(\mu, \sigma)$

Standardavvik

Forventningsverdi

Normalfordelt

Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
- 7. Parameterestimering og intervaller**
8. Hypotesetester
9. Lineær regresjon

Uke 7: Eksempel

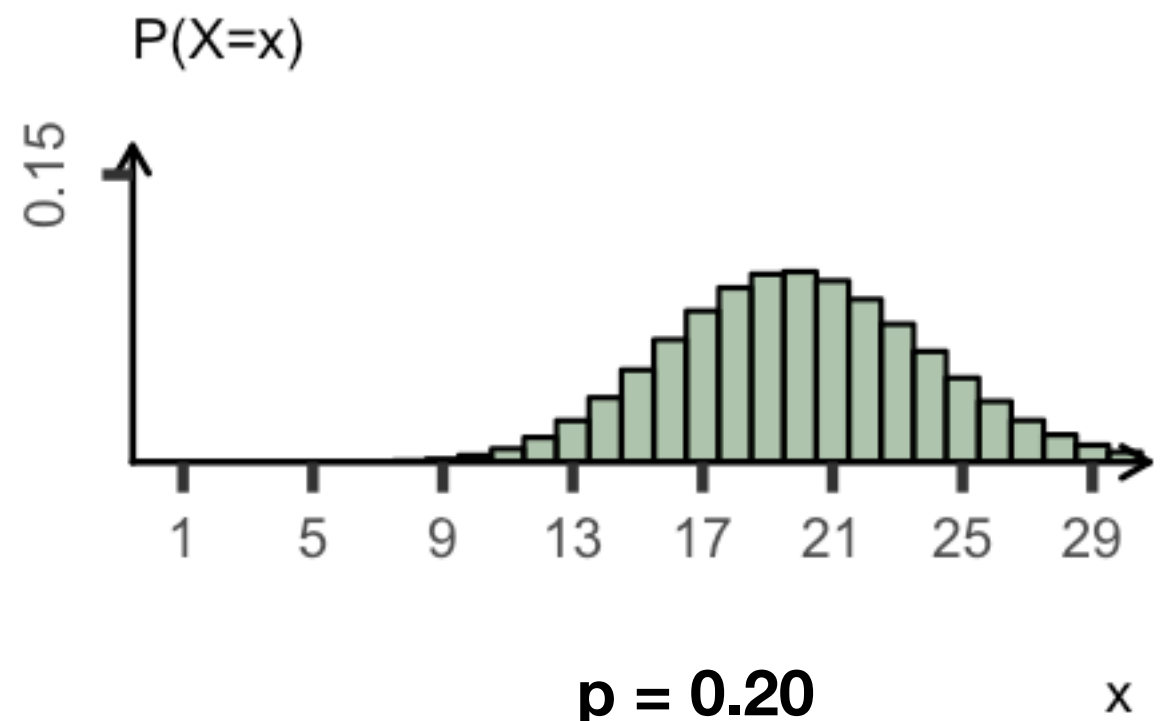
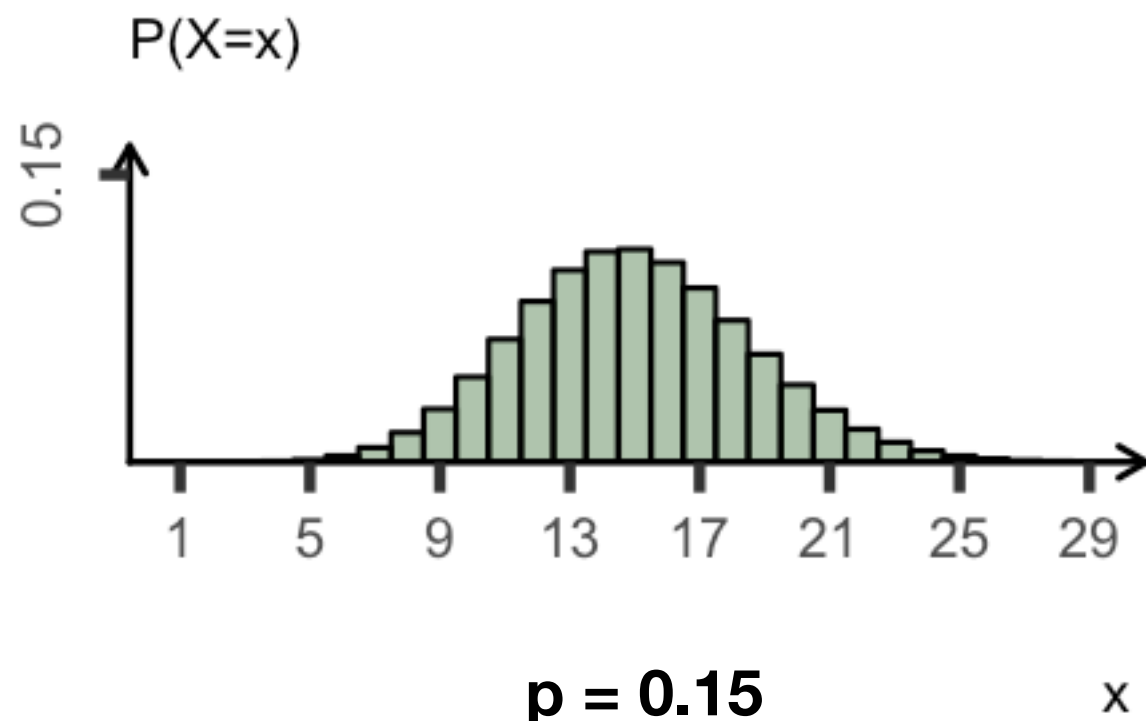
I en befolkning skal det innføres en ny vaksine

Myndighetene antar at en viss andel av befolkningen, p , er skeptiske til dette

Ved å spørre et tilfeldig utvalg personer vil de estimere andelen p

I et utvalg på 100 personer, svarte 17 personer at de er skeptiske

Regn ut et estimat på andelen p og et tilnærmet 95% konfidensintervall



Fellesmodul

1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
- 8. Hypotesetester**
9. Lineær regresjon

Uke 8: Eksempel

En løpesko produsent hevder at deres nye sko kan forbedre din gjennomsnittlige løpehastighet (min/km)

Martin løper i gjennomsnitt 6 min/km, men med noe variasjon

Han antar at gjennomsnittshastighet over én løpetur kan antas å være normalfordelt med forventningsverdi 6 min/km og standardavvik 0.5 min/km

Martin planlegger å løpe 10 (uavhengige) løpeturer *i de nye skoene* og vil teste om han ser en *forbedring* av løpehastigheten sin

- a) Formuler problemet som en hypotesetest (null og alternativ-hypotese)
- b) Skriv ned en testobservator og forkastningsregel ved signifikansnivå 0.05
- c) Etter 10 turer har Martin registrert et gjennomsnitt på 5.8 min/km.
Hva blir utfallet av hypotesetesten?

Fellesmodul

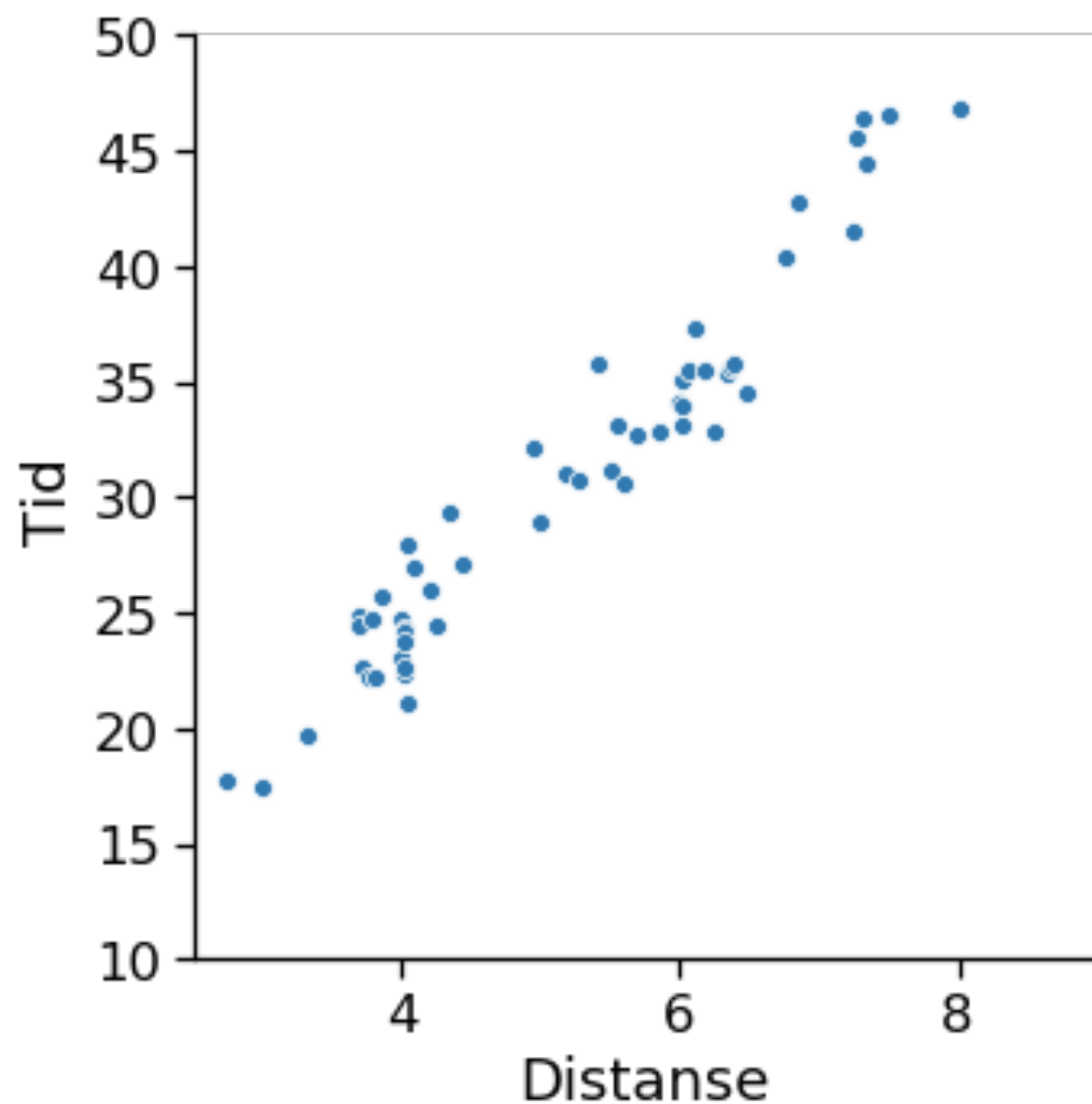
1. Beskrivende statistikk
2. Hendelser og sannsynlighet
3. Stokastiske variabler
4. Binomisk forsøksrekke
5. Poissonprosess
6. Normalfordeling
7. Parameterestimering og intervaller
8. Hypotesetester
9. Lineær regresjon

Eksempel: Treningsdata

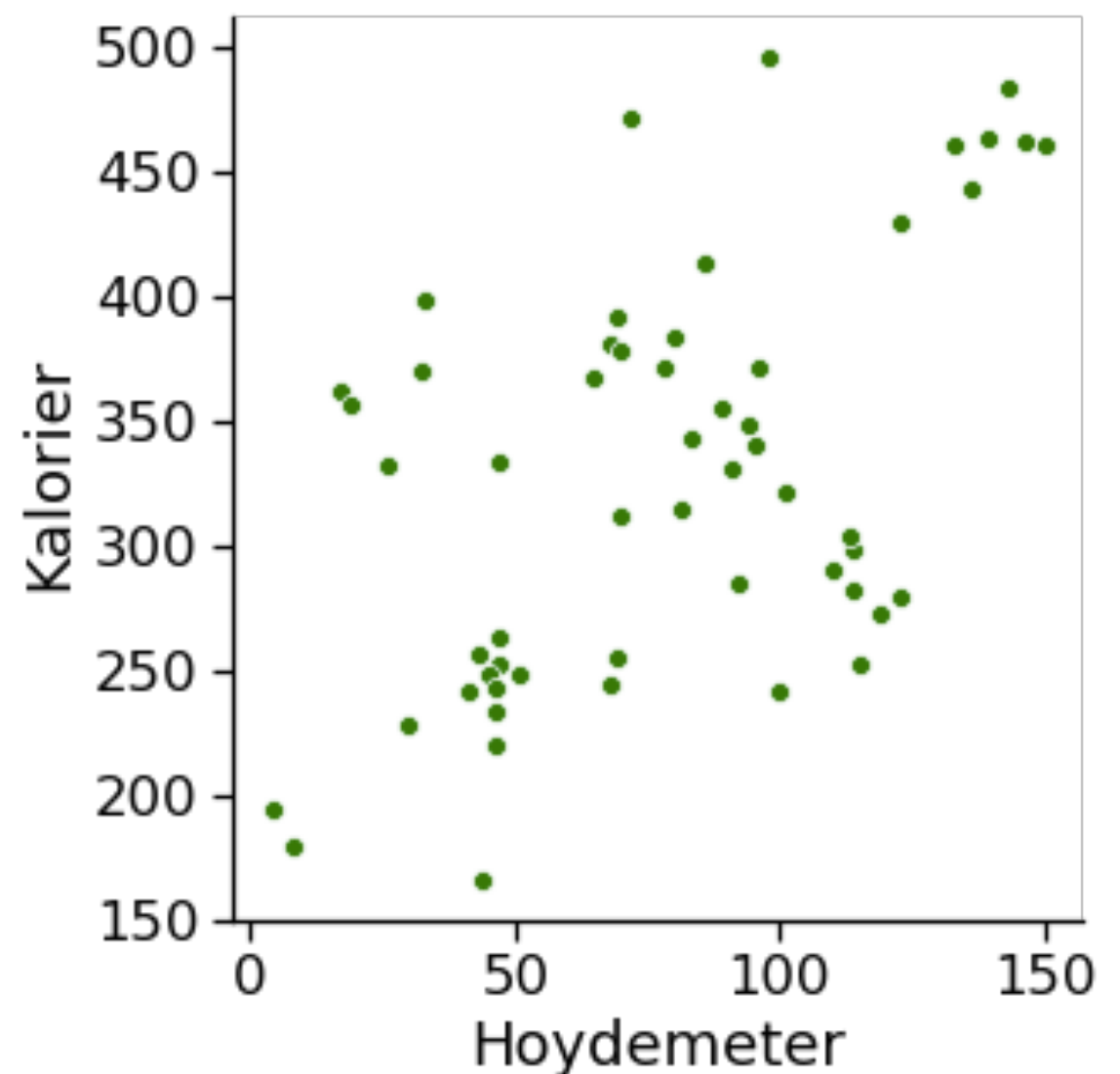


X : distanse (km) Y : tid (minutter)

$(x_1, y_1), (x_2, y_2), \dots, (x_{55}, y_{55})$



Korrelasjon: $R = 0.97$



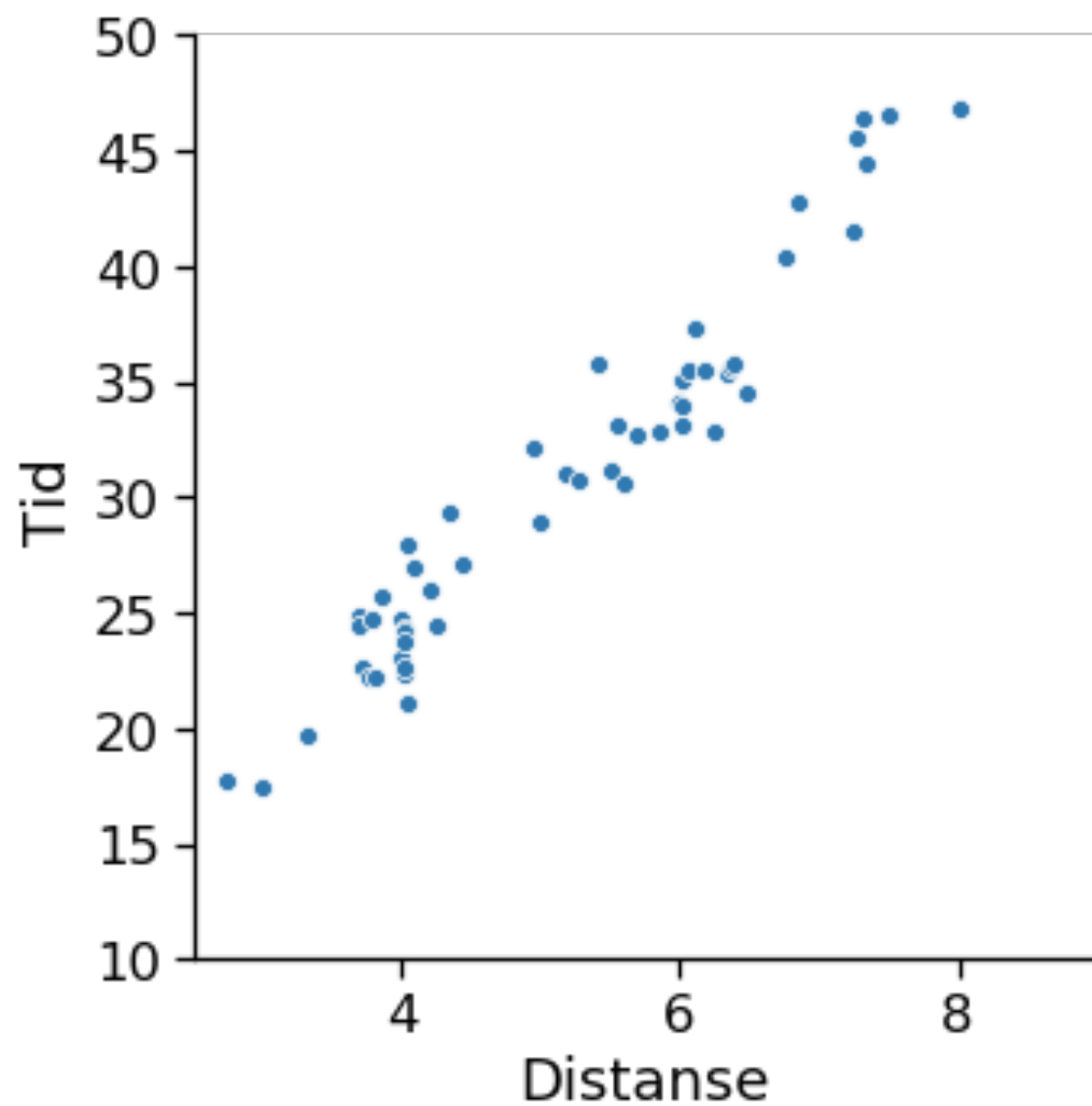
$R = 0.53$

Eksempel: Treningsdata



X : distanse (km) Y : tid (minutter)

$(x_1, y_1), (x_2, y_2), \dots, (x_{55}, y_{55})$

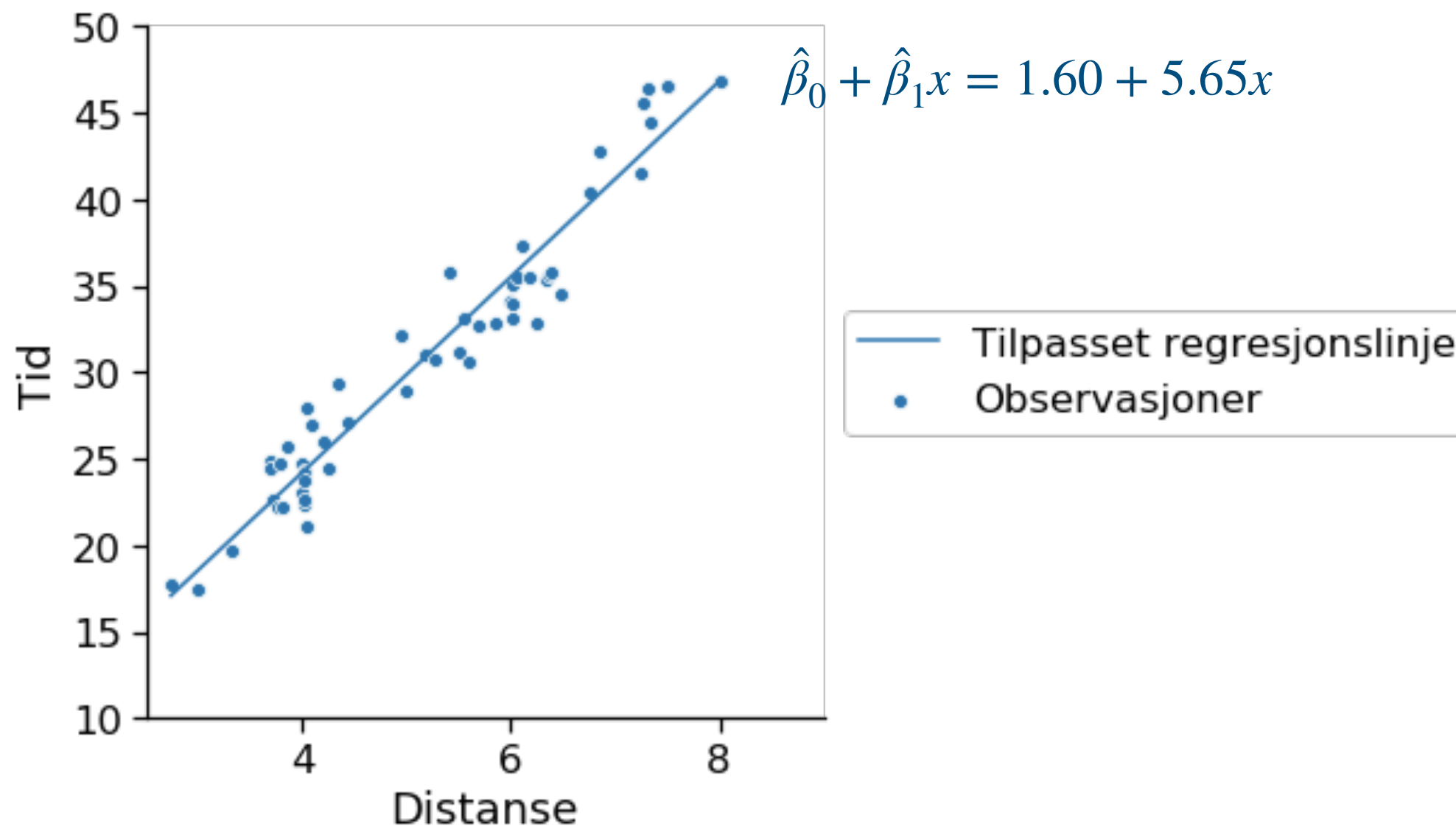


Eksempel: Treningsdata



X : distanse (km) Y : tid (minutter)

$(x_1, y_1), (x_2, y_2), \dots, (x_{55}, y_{55})$



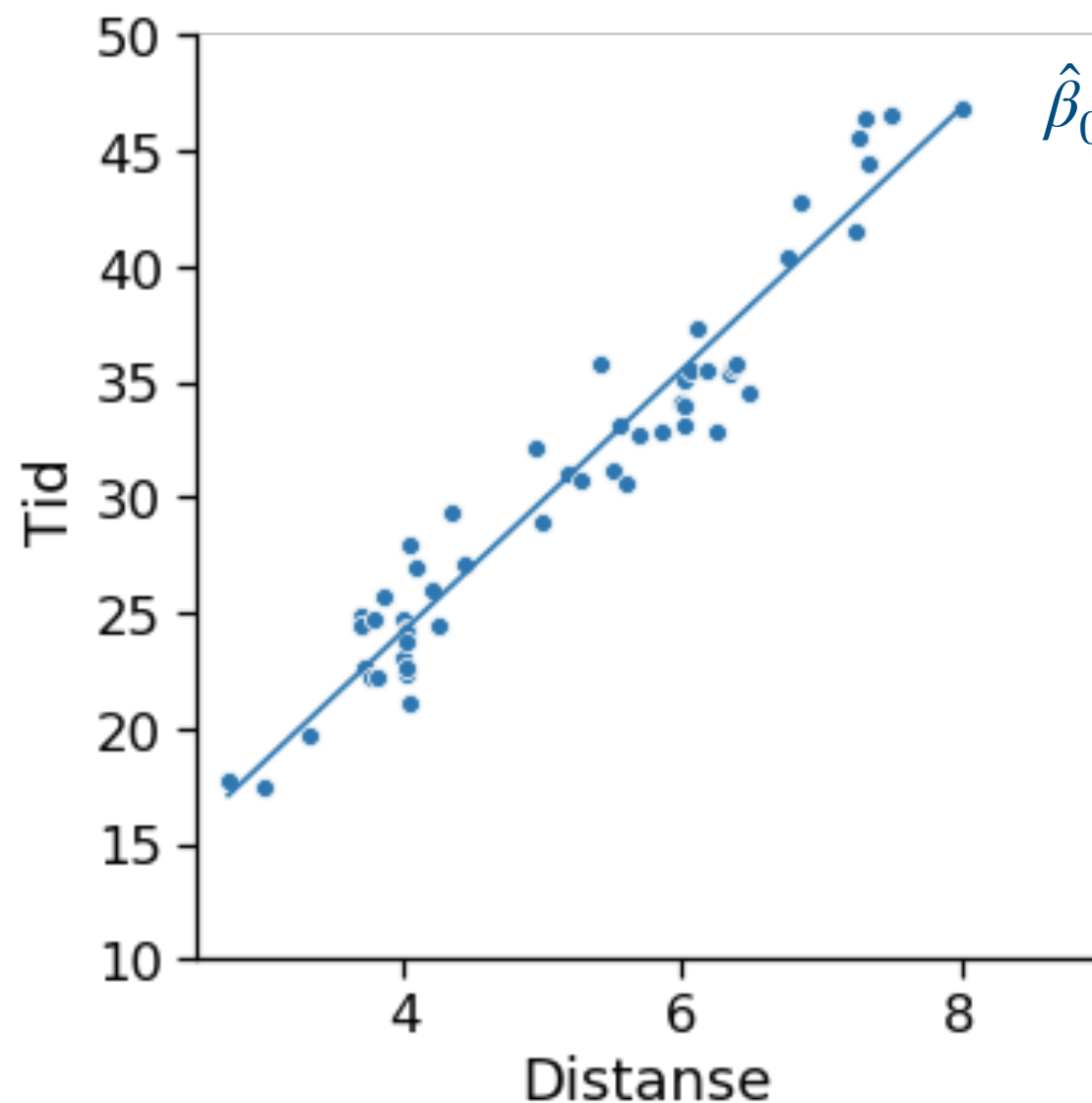
Korrelasjon: $R = 0.97$

Eksempel: Treningsdata



X : distanse (km) Y : tid (minutter)

$(x_1, y_1), (x_2, y_2), \dots, (x_{55}, y_{55})$



$$\hat{\beta}_0 + \hat{\beta}_1 x = 1.60 + 5.65x$$

95% KI for β_1 : [5.25, 6.05]

— Tilpasset regresjonslinje
• Observasjoner

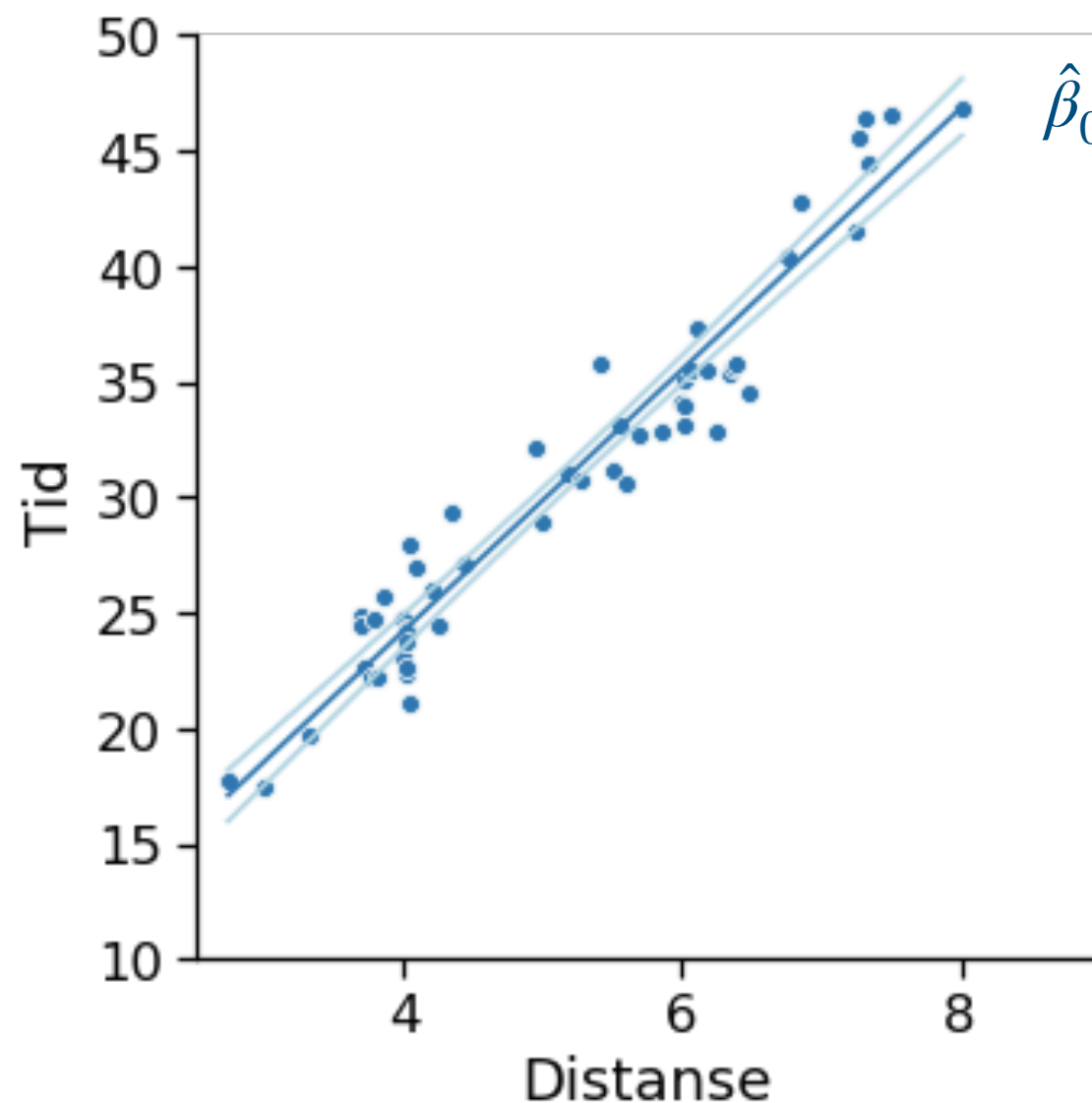
Korrelasjon: $R = 0.97$

Eksempel: Treningsdata



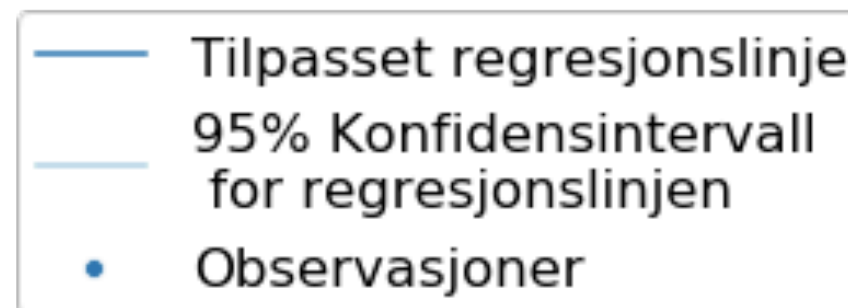
X : distanse (km) Y : tid (minutter)

$(x_1, y_1), (x_2, y_2), \dots, (x_{55}, y_{55})$



$$\hat{\beta}_0 + \hat{\beta}_1 x = 1.60 + 5.65x$$

95% KI for β_1 : [5.25, 6.05]

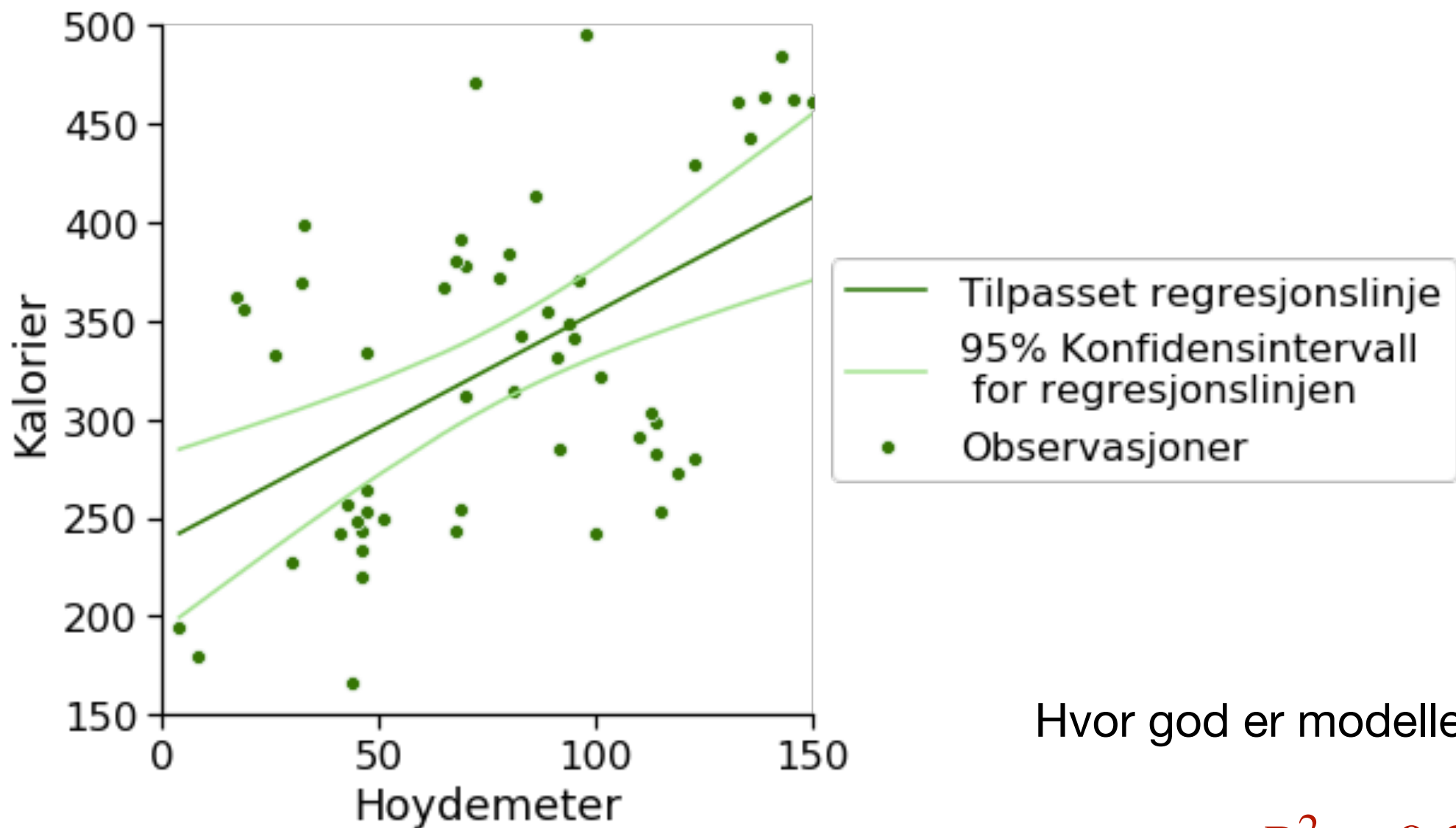


Hvor god er modellen?

$$R^2 = 0.94$$

Korrelasjon: $R = 0.97$

Eksempel: Treningsdata



Hvor god er modellen?

$$R^2 = 0.28$$

Korrelasjon: $R = 0.53$