

Lineær regresjon

Del 4

Thea Bjørnland
Institutt for matematiske fag
NTNU

Lineær regresjon: Modellantagelser

Mål: $Y|X = x$

Antagelse 1: Lineær sammenheng

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

Antagelse 2: Y , betinget på $X=x$, er normalfordelt med standardavvik σ uansett x

$$Y|X = x \sim N(\beta_0 + \beta_1 x, \sigma)$$

Alternativ formulering:

$$Y = \underbrace{\beta_0 + \beta_1 x}_{\text{Ikke stokastisk}} + e$$

Ikke stokastisk

$$e \sim N(0, \sigma)$$

Stokastisk

Feil-ledd

Hvordan responsen Y
tar verdier rundt regresjonslinja

Lineær regresjon: Modellantagelser

Tilfeldig utvalg: $(x_1, Y_1), \dots, (x_n, Y_n)$ OBS: Antar uavhengige variabler!

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim N(0, \sigma) \quad i = 1, \dots, n$$

(uavhengige feilledd)

Observasjoner: $(y_1, x_1), \dots, (y_n, x_n)$ Estimert regresjonslinje: $\hat{\beta}_0 + \hat{\beta}_1 x_i$

Residualer: $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Lineær regresjon: Modellantagelser

Tilfeldig utvalg: $(x_1, Y_1), \dots, (x_n, Y_n)$ **OBS: Antar uavhengige variabler!**

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim N(0, \sigma) \quad i = 1, \dots, n$$

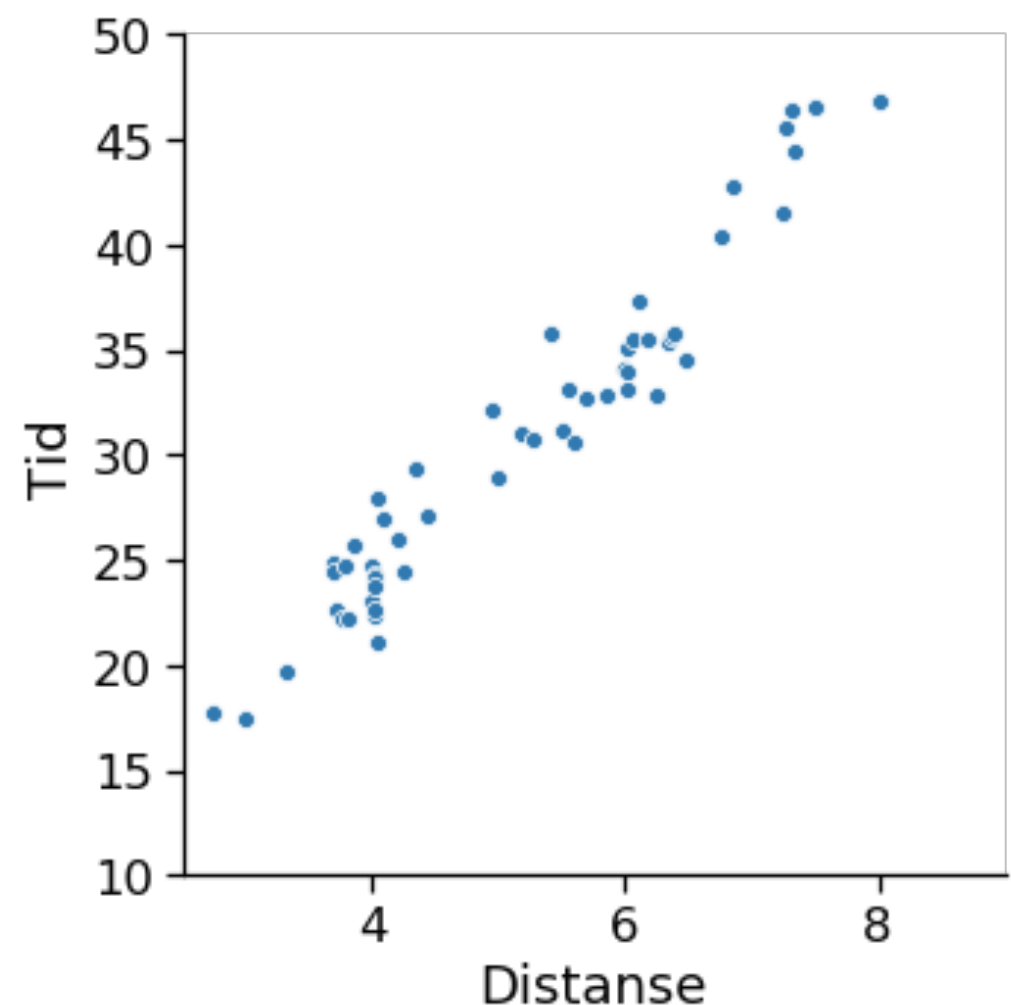
(uavhengige feilledd)

Observasjoner: $(y_1, x_1), \dots, (y_n, x_n)$ Estimert regresjonslinje: $\hat{\beta}_0 + \hat{\beta}_1 x_i$

Residualer: $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Modellsjekk: Treningsdata

Antagelse: Lineær sammenheng ✓



Lineær regresjon: Modellantagelser

Tilfeldig utvalg: $(x_1, Y_1), \dots, (x_n, Y_n)$ OBS: Antar uavhengige variabler!

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim N(0, \sigma) \quad i = 1, \dots, n$$

(uavhengige feilledd)

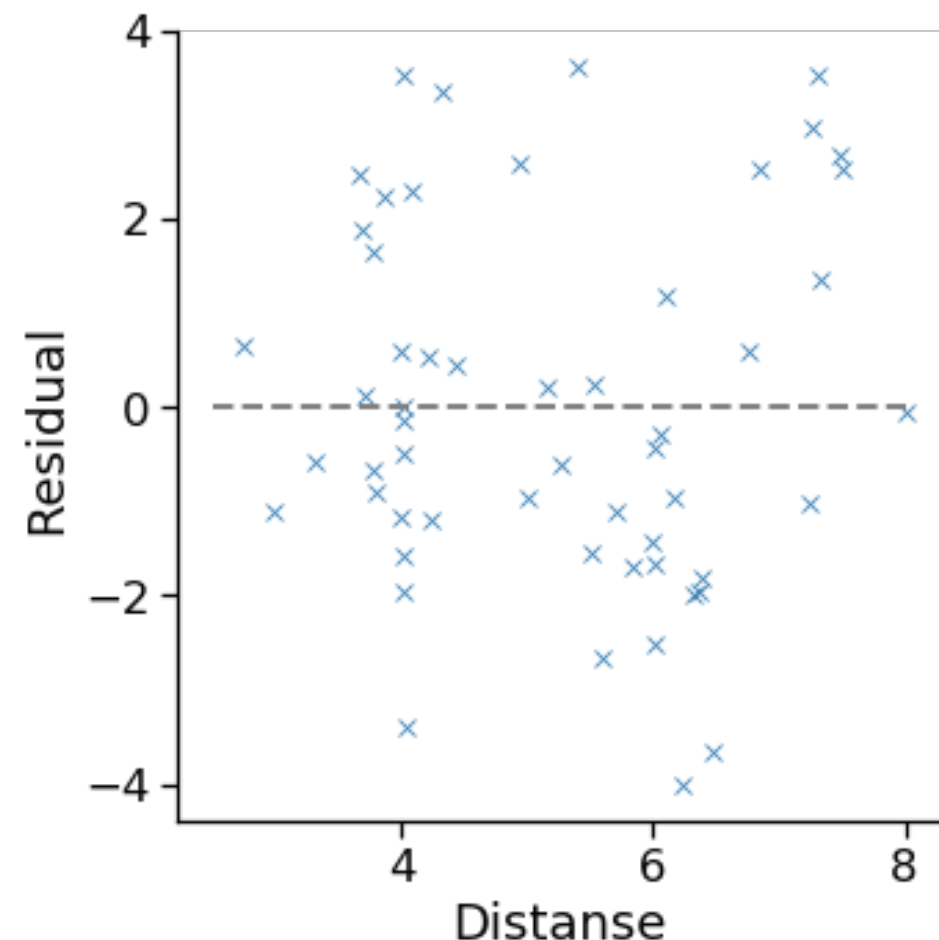
Observasjoner: $(y_1, x_1), \dots, (y_n, x_n)$ Estimert regresjonslinje: $\hat{\beta}_0 + \hat{\beta}_1 x_i$

Residualer: $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Modellsjekk: Treningsdata

Antagelse: Lineær sammenheng ✓

Antagelse: $e_i \sim N(0, \sigma)$



Lineær regresjon: Modellantagelser

Tilfeldig utvalg: $(x_1, Y_1), \dots, (x_n, Y_n)$ **OBS: Antar uavhengige variabler!**

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim N(0, \sigma) \quad i = 1, \dots, n$$

(uavhengige feilledd)

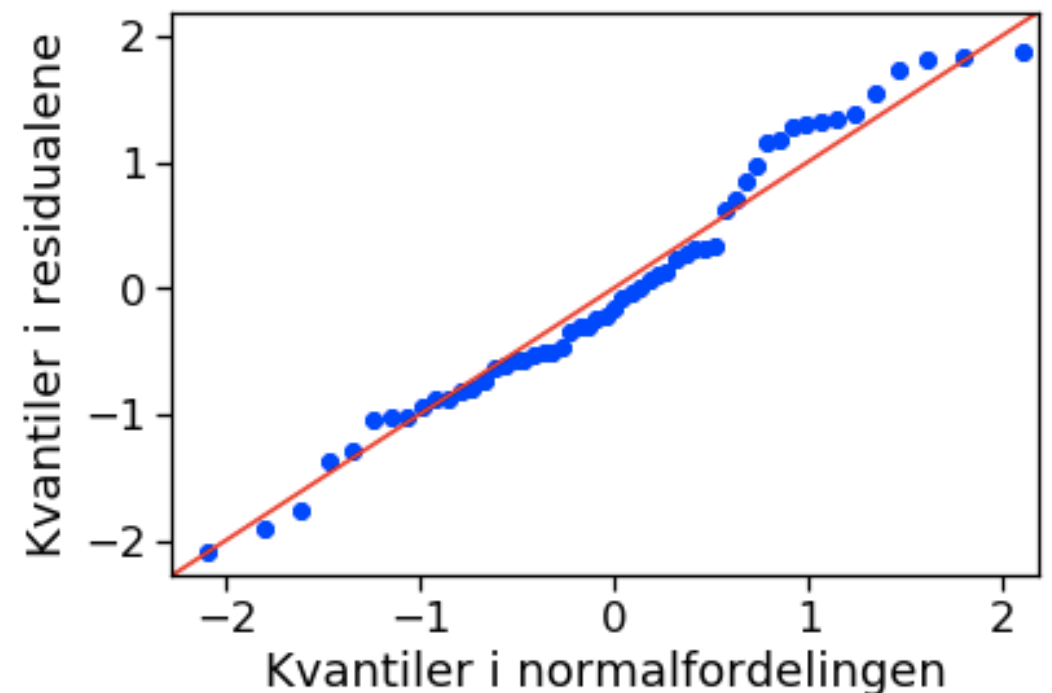
Observasjoner: $(y_1, x_1), \dots, (y_n, x_n)$ Estimert regresjonslinje: $\hat{\beta}_0 + \hat{\beta}_1 x_i$

Residualer: $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Modellsjekk: Treningsdata

Antagelse: Lineær sammenheng ✓

Antagelse: $e_i \sim N(0, \sigma)$



Lineær regresjon: Modellantagelser

Tilfeldig utvalg: $(x_1, Y_1), \dots, (x_n, Y_n)$ OBS: Antar uavhengige variabler!

$$Y_i = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim N(0, \sigma) \quad i = 1, \dots, n$$

(uavhengige feilledd)

Observasjoner: $(y_1, x_1), \dots, (y_n, x_n)$ Estimert regresjonslinje: $\hat{\beta}_0 + \hat{\beta}_1 x_i$

Residualer: $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Modellsjekk: Treningsdata

Antagelse: Lineær sammenheng ✓

Antagelse: $e_i \sim N(0, \sigma)$
✓ Uavhengige? ✓

