

Enkel lineær regresjon: vurdering av modellantagelsene

- ▶ Modellantagelser i enkel lineær regresjon:

- i)* $E[Y_i] = \alpha + \beta x_i$

- ii)* $\text{Var}[Y_i] = \sigma^2$

- iii)* Y_i er normalfordelt

- iv)* Y_1, Y_2, \dots, Y_n er uavhengige

- ▶ For et gitt datasett, hvordan kan vi sjekke om dette synes rimelig?

Enkel lineær regresjon: vurdering av modellantagelsene

- ▶ Modellantagelser i enkel lineær regresjon:
 - i)* $E[Y_i] = \alpha + \beta x_i$
 - ii)* $\text{Var}[Y_i] = \sigma^2$
 - iii)* Y_i er normalfordelt
 - iv)* Y_1, Y_2, \dots, Y_n er uavhengige
- ▶ For et gitt datasett, hvordan kan vi sjekke om dette synes rimelig?
 - ▶ se på kryssplott (*i*)

Enkel lineær regresjon: vurdering av modellantagelsene

- ▶ Modellantagelser i enkel lineær regresjon:
 - i)* $E[Y_i] = \alpha + \beta x_i$
 - ii)* $\text{Var}[Y_i] = \sigma^2$
 - iii)* Y_i er normalfordelt
 - iv)* Y_1, Y_2, \dots, Y_n er uavhengige
- ▶ For et gitt datasett, hvordan kan vi sjekke om dette synes rimelig?
 - ▶ se på kryssplott (*i*)
 - ▶ ut fra hvordan observasjonene er gjort/forsøksopplegget (*iv*)

Enkel lineær regresjon: vurdering av modellantagelsene

- ▶ Modellantagelser i enkel lineær regresjon:

i) $E[Y_i] = \alpha + \beta x_i$

ii) $\text{Var}[Y_i] = \sigma^2$

iii) Y_i er normalfordelt

iv) Y_1, Y_2, \dots, Y_n er uavhengige

- ▶ For et gitt datasett, hvordan kan vi sjekke om dette synes rimelig?

- ▶ se på kryssplott (*i*)

- ▶ ut fra hvordan observasjonene er gjort/forsøksopplegget (*iv*)

- ▶ se på qq-plott: (*iii*)

- ▶ modellen er: $Y_i = \alpha + \beta x_i + \varepsilon_i \Leftrightarrow \varepsilon_i = Y_i - \alpha - \beta x_i$

- ▶ estimerte residualer: $\hat{\varepsilon}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$

- ▶ qq-plot for $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n$

Enkel lineær regresjon: vurdering av modellantagelsene

- ▶ Modellantagelser i enkel lineær regresjon:

i) $E[Y_i] = \alpha + \beta x_i$

ii) $\text{Var}[Y_i] = \sigma^2$

iii) Y_i er normalfordelt

iv) Y_1, Y_2, \dots, Y_n er uavhengige

- ▶ For et gitt datasett, hvordan kan vi sjekke om dette synes rimelig?

- ▶ se på kryssplott (*i*)

- ▶ ut fra hvordan observasjonene er gjort/forsøksopplegget (*iv*)

- ▶ se på qq-plott: (*iii*)

- ▶ modellen er: $Y_i = \alpha + \beta x_i + \varepsilon_i \Leftrightarrow \varepsilon_i = Y_i - \alpha - \beta x_i$

- ▶ estimerte residualer: $\hat{\varepsilon}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$

- ▶ qq-plot for $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n$

- ▶ se på residualplott: (*i*, *ii* og *iii*)

- ▶ residualplott: plotte x_i mot $\hat{\varepsilon}_i$

- ▶ vi skal ikke kunne se noen strukturer her

- ▶ merk: vi vil alltid ha $\sum_{i=1}^n \hat{\varepsilon}_i = 0$