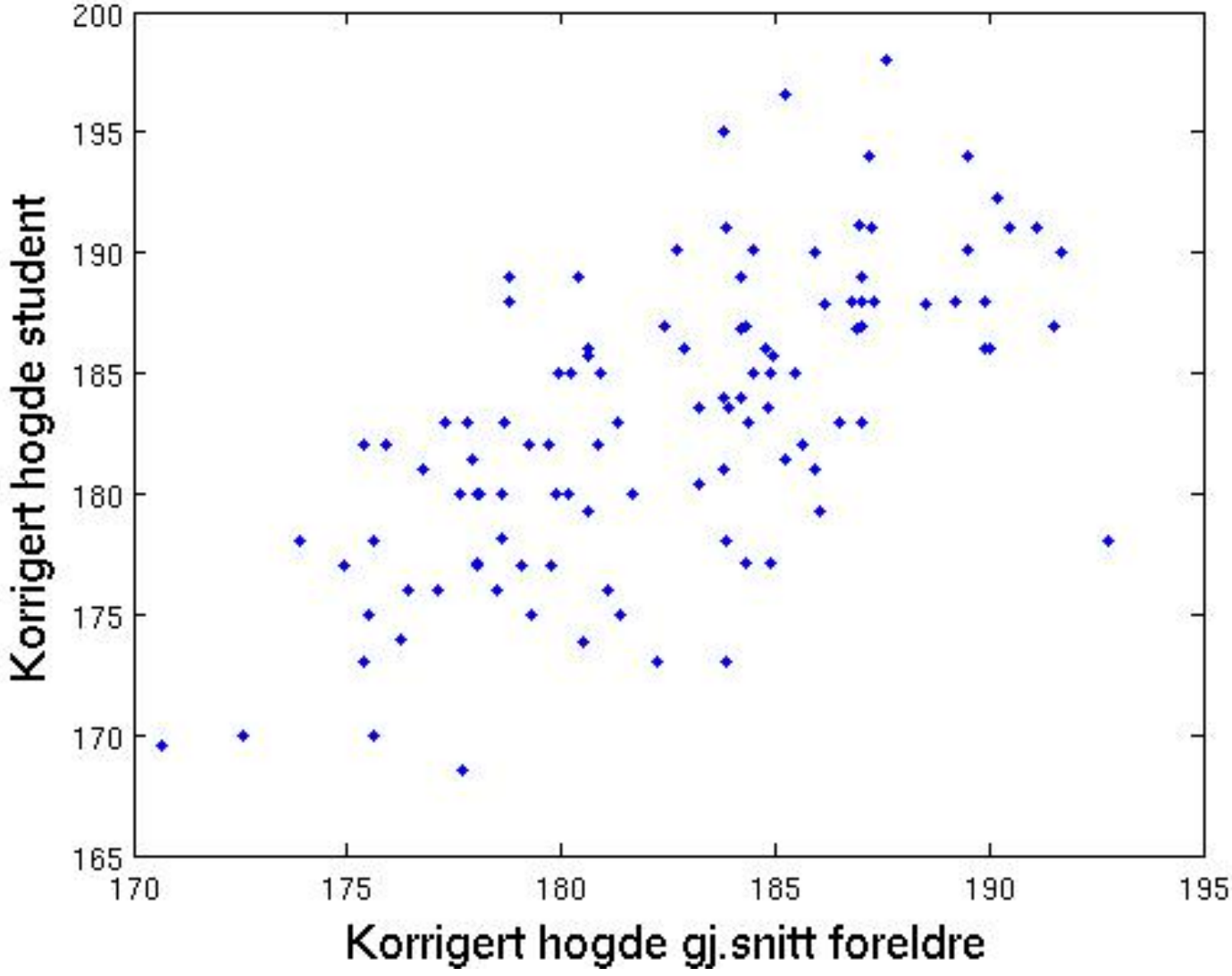
Våre høgdedata



Enkel lineær regresjon

$$Y_i = \alpha + \beta x_i + \epsilon$$

- Y: Respons (stok. var)
- x_i: Forklaringsvariabel (kjent, tal)
- α og β: Regresjonsparameter (param, tal, ukjent)
- ϵ: 'støyen', 'feilen', stok.var
  - $E(\epsilon) = 0$, $Var(\epsilon) = \sigma_\epsilon^2$
  - σ_ϵ^2 (param., tal, ukjent)

Dersom $\epsilon \sim N(0, \sigma_\epsilon^2)$

$$Y_i \sim N(\alpha + \beta x_i, \sigma_\epsilon^2)$$

Regresjonsparametra

Må estimere α, β, σ_ϵ^2 frå data.
Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Estimatorar

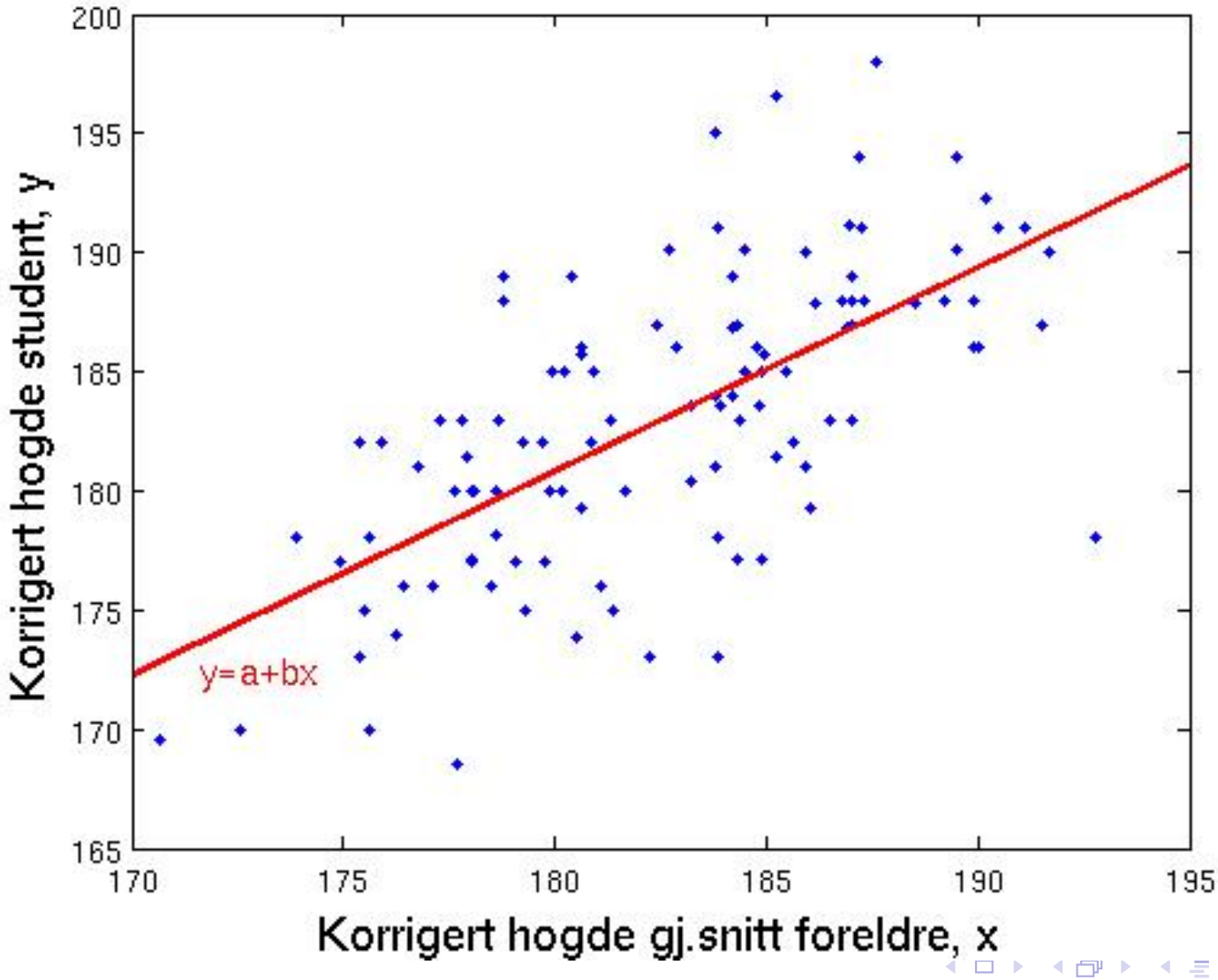
- $A = \bar{Y} - B\bar{x}$
- $B = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- $S_\epsilon^2 = \dfrac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}$
  - Der $\hat{Y}_i = A + Bx_i$

Estimat

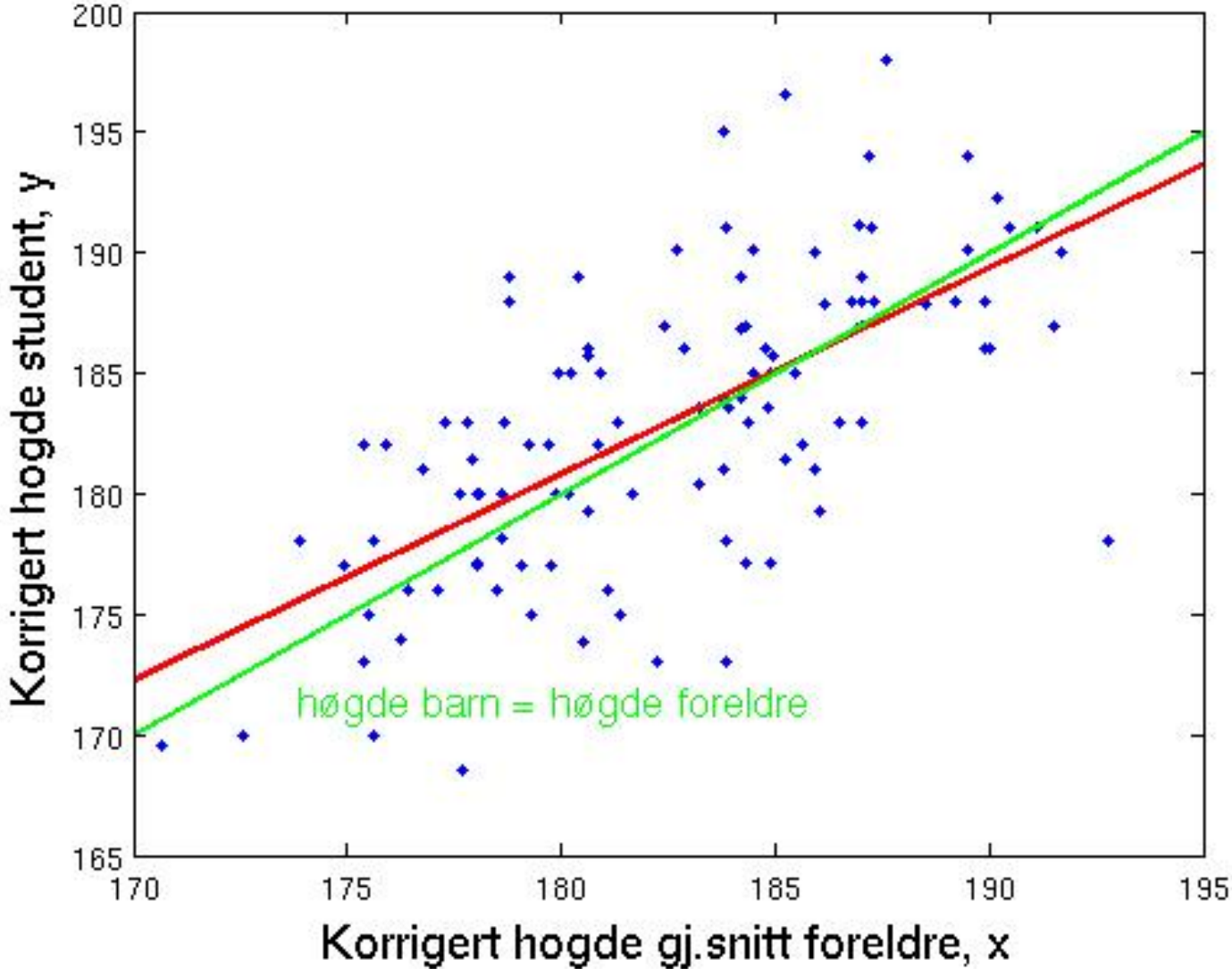
- Stok.var $\longrightarrow$ data
- a, b, s_ϵ^2, $\hat{y}$

Regresjonslinje (95% KI), våre data $n = 110$

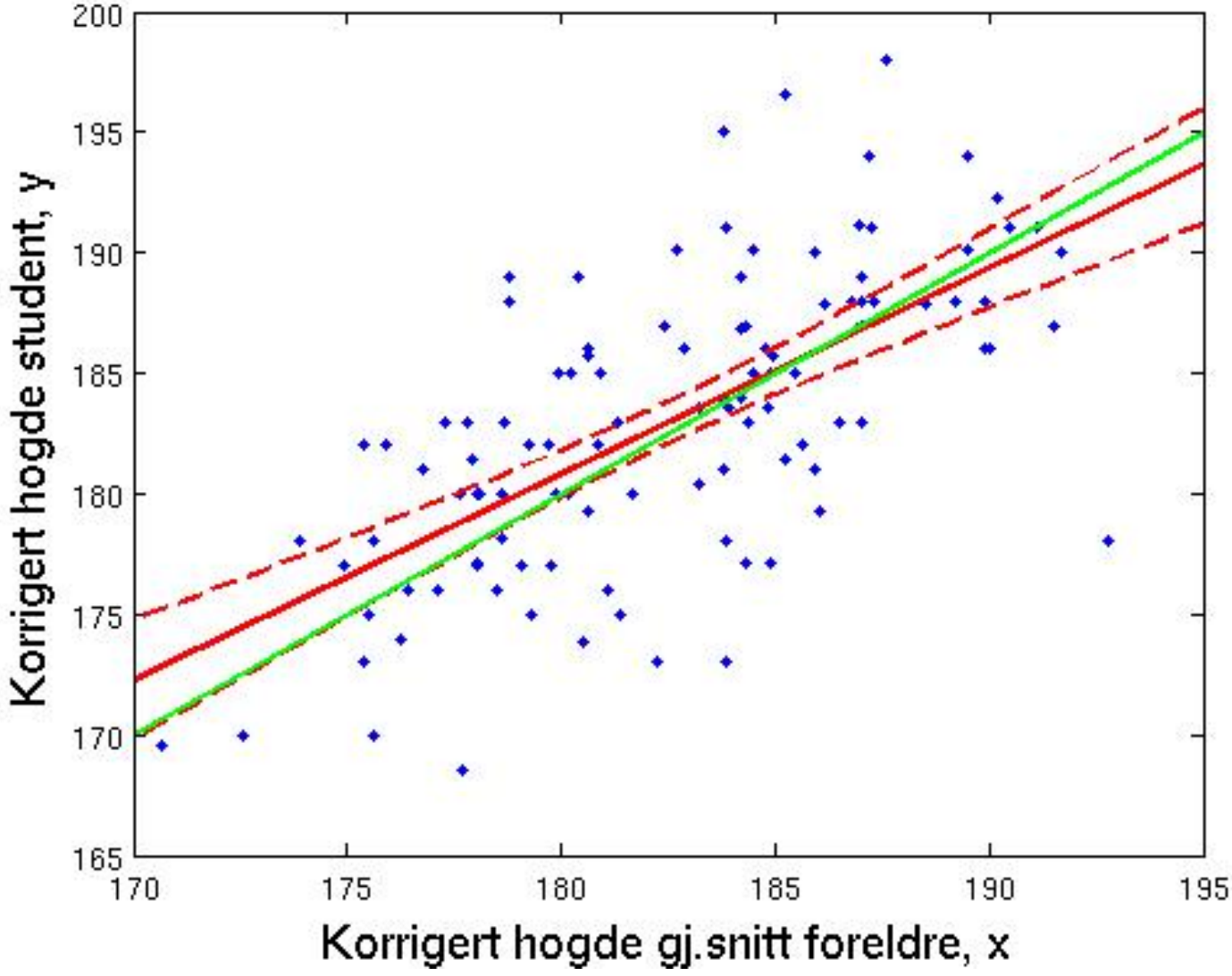
Estimat: $a = 27.2$ $(-6.6451, 60.9989)$, $b = 0.85$ $(0.6684, 1.0387)$ og $s^2 = 4.64^2$.



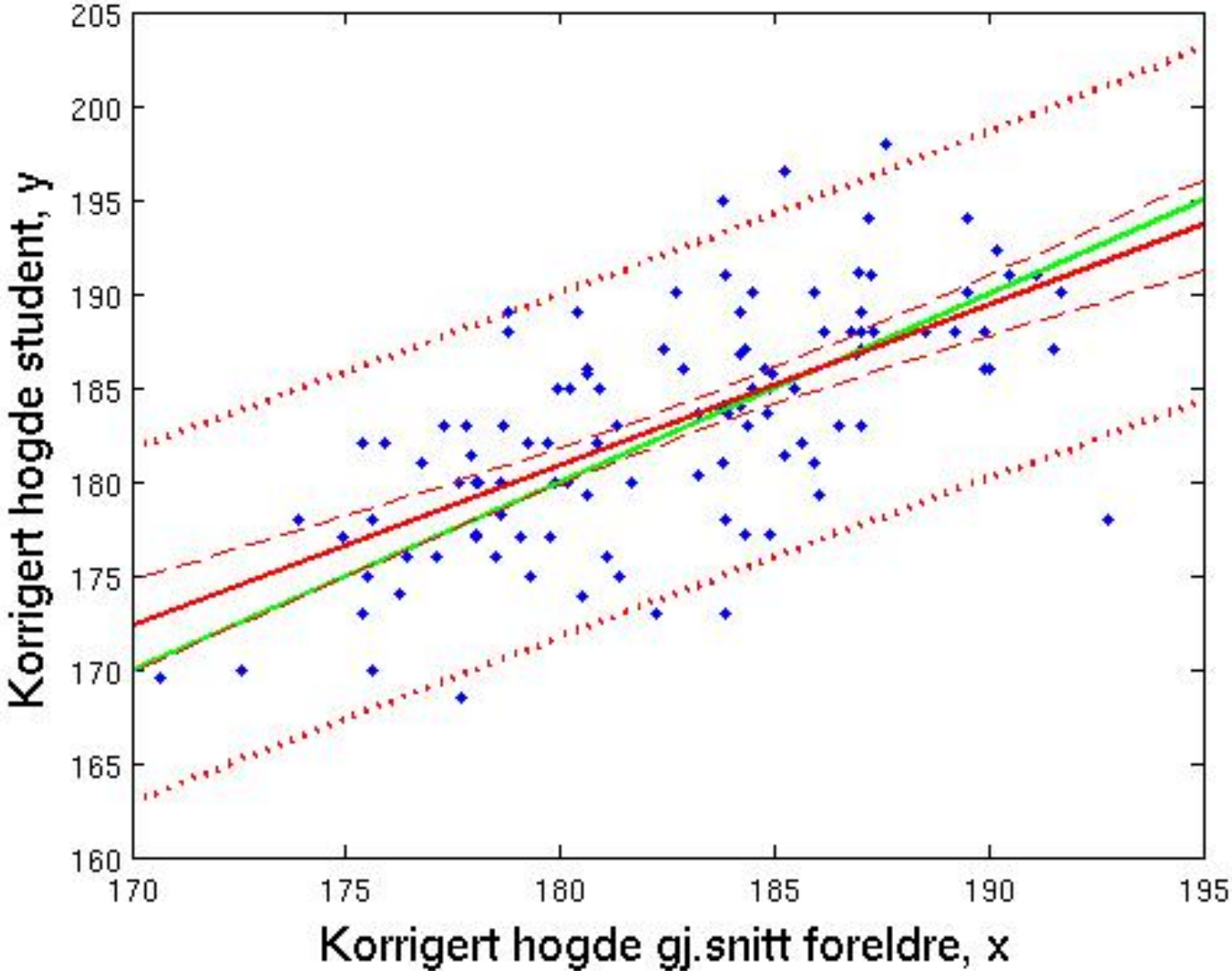
Høgde barn = høgde foreldre?



95% konfidensintervall



95% prediksjonsintervall



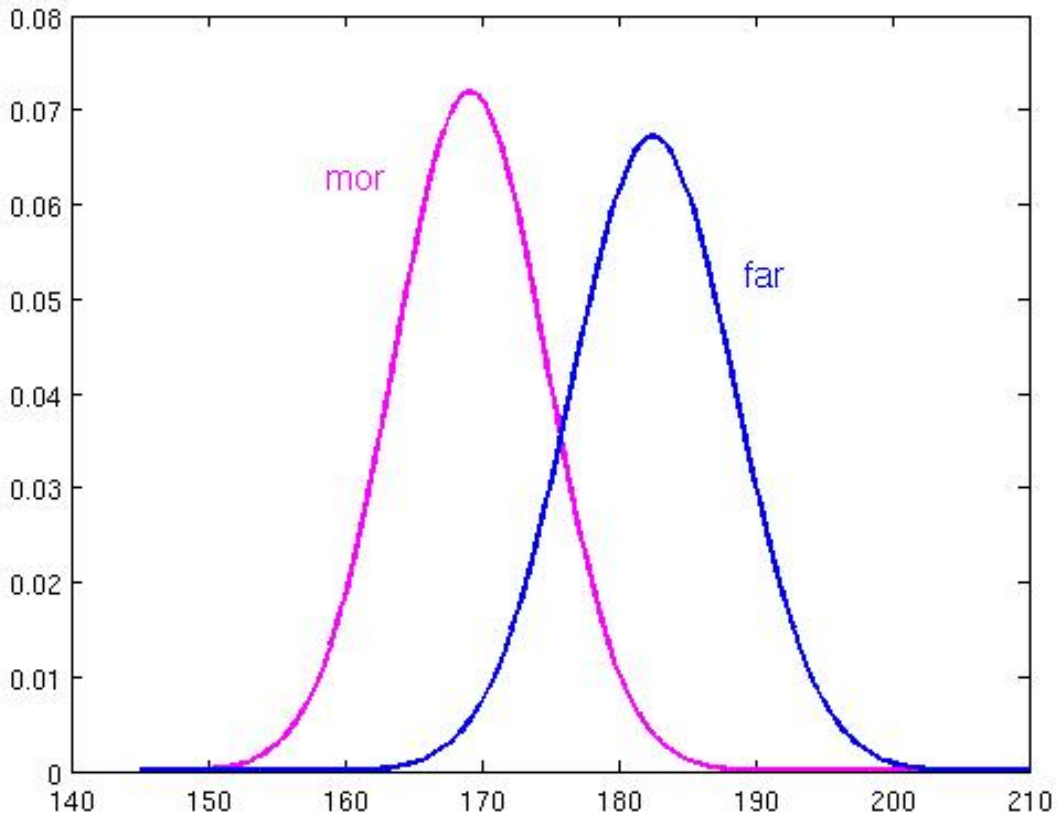
Kap. 11.12 Korrelasjon

Skal binde saman:

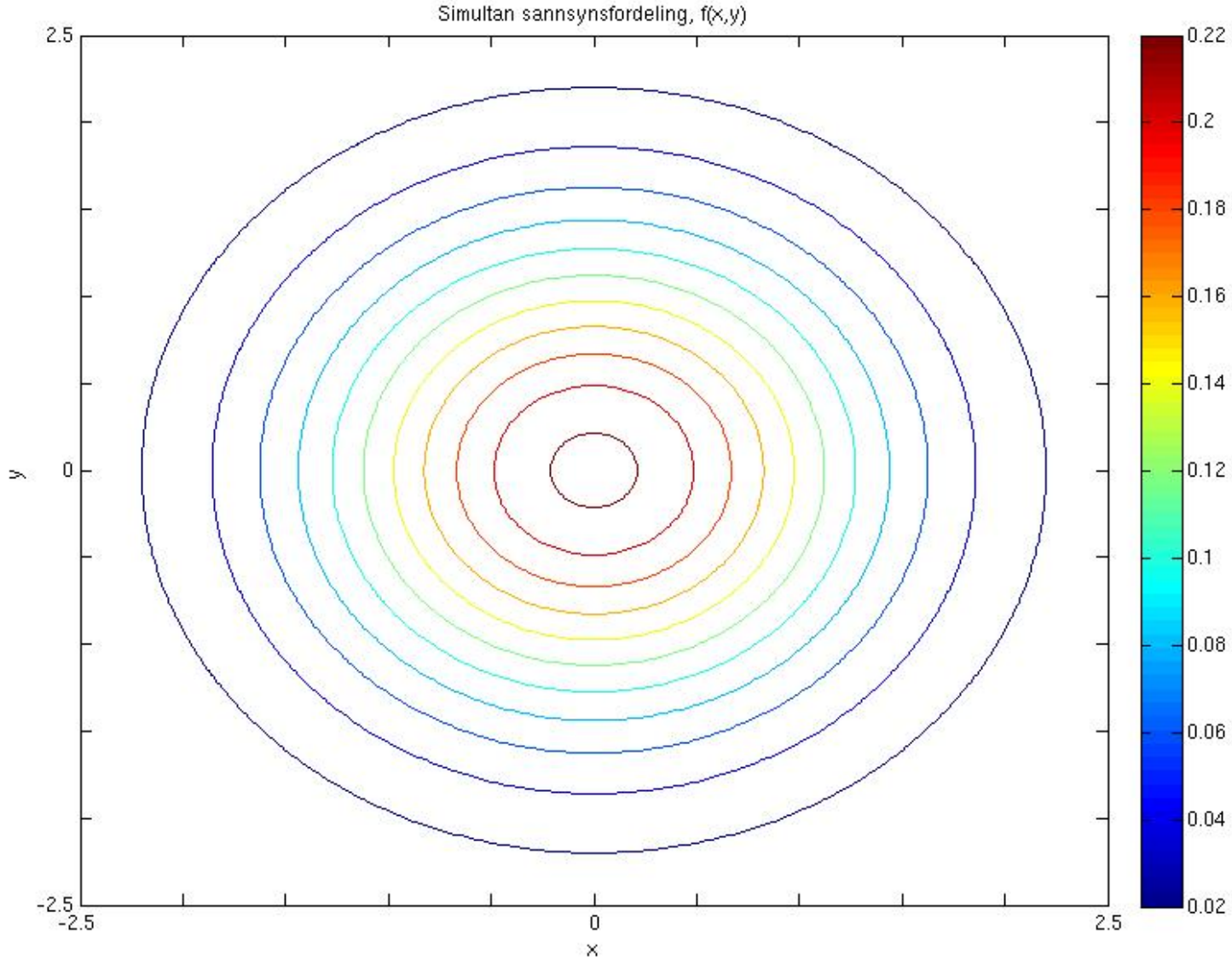
- Kap. 3.4 Simultan sannsynsfordeling
  - To stok.var., X og Y, $f(x, y)$
- Kap. 4.2 Varians og kovarians av stokastiske variable
  - Kovarians $\sigma_{XY} = E((X - \mu_X)(Y - \mu_Y))$
  - Korrelasjon $\rho = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$, $-1 < \rho < 1$
- Kap. 11.2 Enkel lineær regresjon
  - Kjenner forklaringsvariabelen x. $Y = \alpha + \beta x + \epsilon$, $\epsilon \sim N(0, \sigma_\epsilon^2)$.
- Kap. 11.5 .... R^2: Tilpassingskoeffisienten.
  - $R^2 = 1 - \dfrac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$.

    e_i: Residual

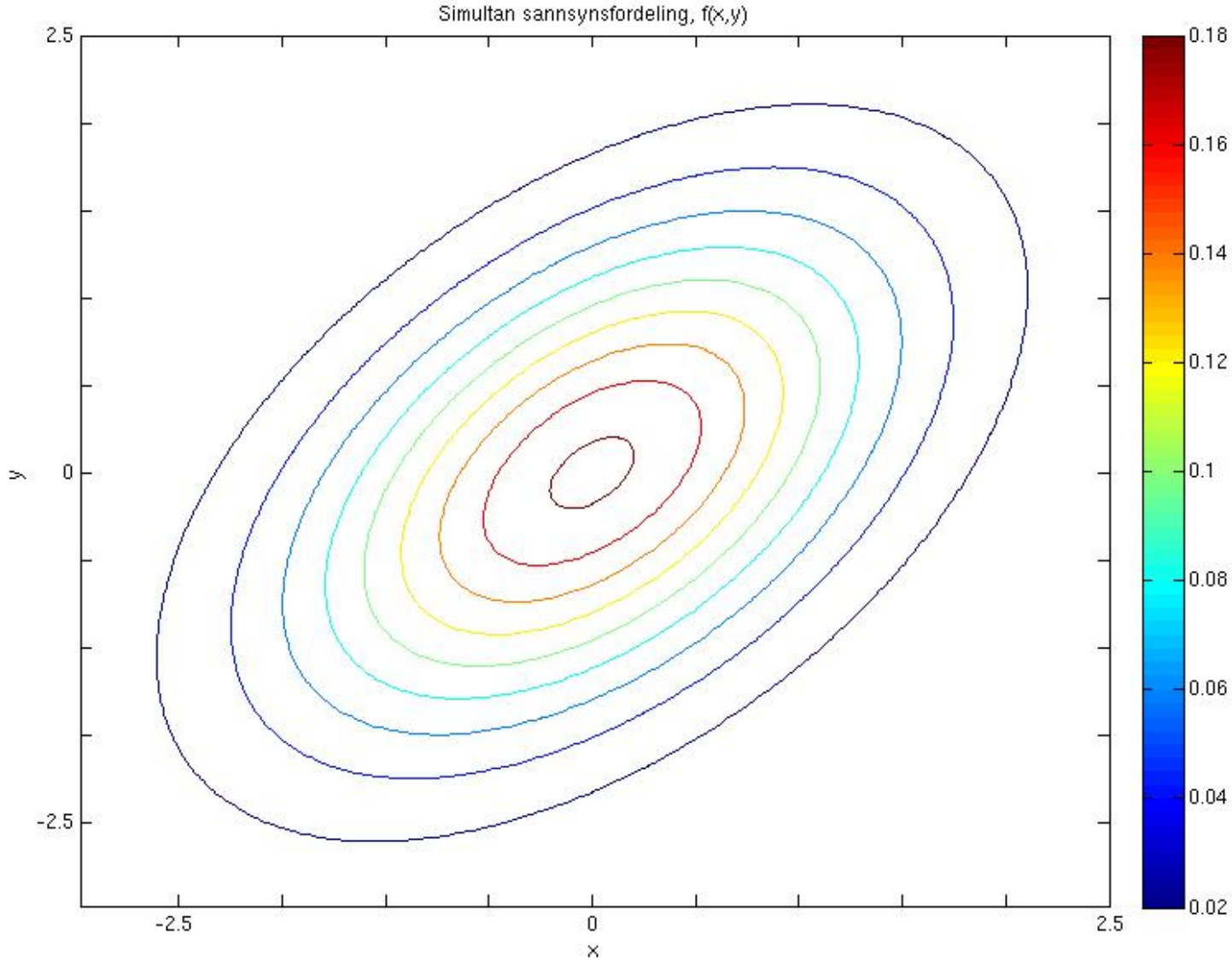
Mor - far marginale sanns.fordeling og data



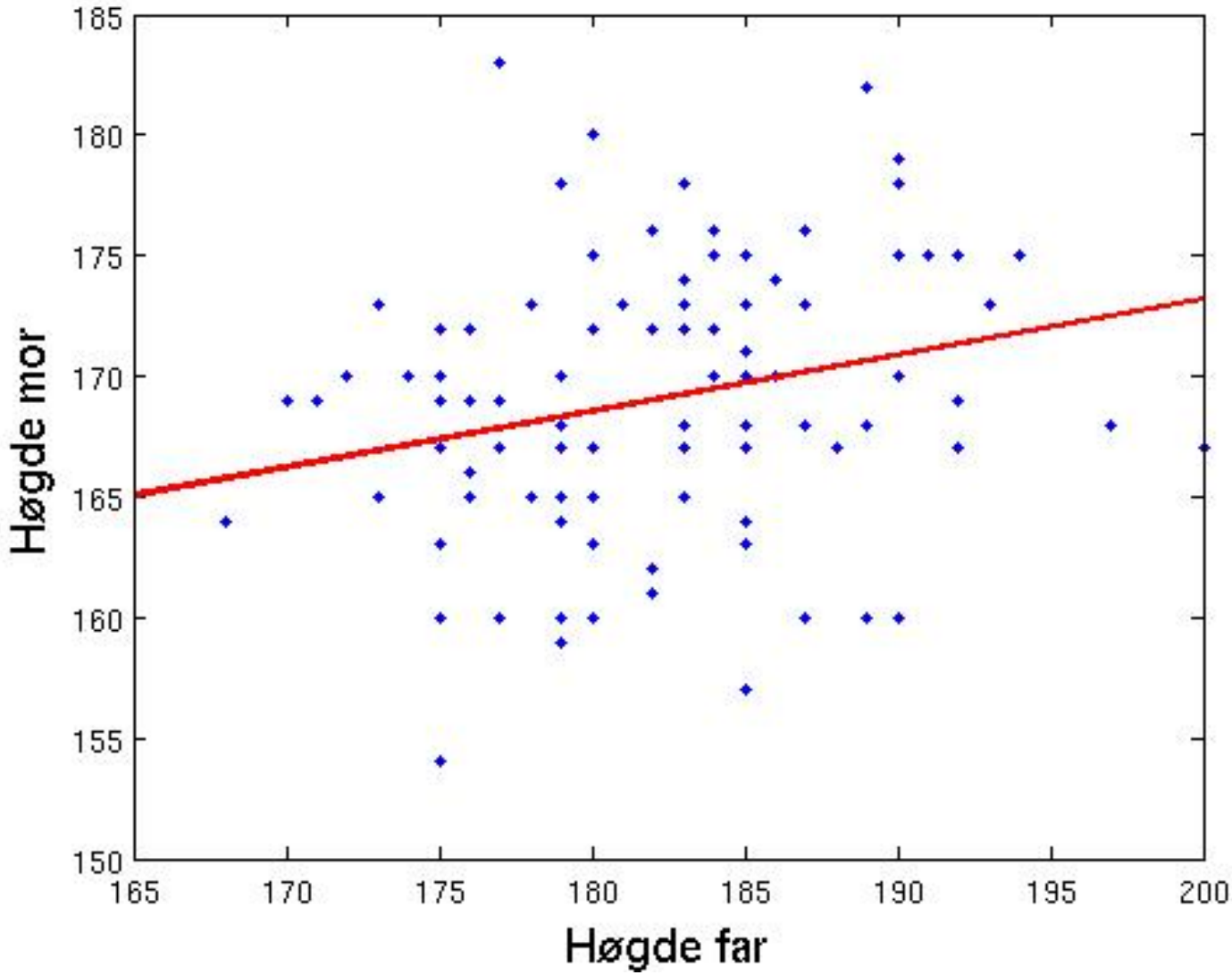
Simultanfordeling 1, marginale N(0,1)



Simultanfordeling 2, marginale N(0,1)



Lineær modell høgde far høgde mor



Simultanfordeling høgde far høgde mor