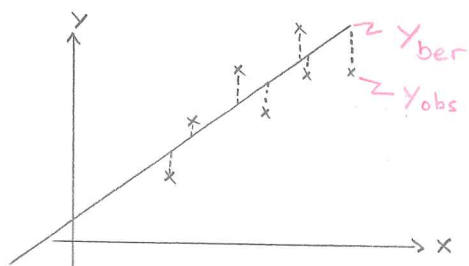


Regression används för att finna bästa värde på justerbara parametrar i en modell. Med bästa avses värden som åter speglar mätdata så bra som möjligt. Hur säker bestämningen är kan beskrivas med konfidensintervall, som är vanligt förekommande i statistik.

Ex Modell: $y = ax + b$



Definiera residualen $\Gamma_i = Y_{obs,i} - Y_{ber,i}$ som beskriver skillnad mellan observerat värde och beräknat värde i mätpunkt i .

För att minimera skillnaden vill vi minimera kvadratsumman SS enligt:

$$SS = \min_{a,b} \sum_i \Gamma_i^2$$

Hur blir det för icke linjära modeller i regressions sammanhang?

Ex Modell: $y = a + bx + cx^2$ (icke linjär)

En godtycklig mätpunkt $Y_{obs,i}$ kan beskrivas av

$$Y_{obs,i} = a + bx_i + cx_i^2 + \varepsilon_i, \text{ där } \varepsilon_i \text{ är avvikelserna från modellen.}$$

På matrisform: $Y = X\beta + \varepsilon = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

Linjär regression kan användas om modellen kan skrivas som linjär med avseende på parametrarna.

Finns det någon enklare metod för regressionen än att minimera kvadratsumman med optimering? För ett minimum är

$$\frac{\partial SS}{\partial \text{Parameter}} = 0, \quad SS = \sum_i (y_{\text{obs}} - a - bx - \dots)^2$$

$$\frac{\partial SS}{\partial a} = 0 \Rightarrow 2 \sum_i - () = 0$$

$$\frac{\partial SS}{\partial b} = 0 \Rightarrow 2 \sum_i -x () = 0$$

⋮

Kan bestämma parametrarna
a, b, ...

Allmänt: $X\beta = Y \Rightarrow X^T X \beta = X^T Y$

$$\Rightarrow \beta = X^T (X^T X)^{-1} Y$$

a
b
⋮

Bestäms av mätdata
mätpunkter

Hur blir det om modellen inte kan skrivas som linjär med avseende på parametrarna? Ex: $y = a/(b-x^c)$

$$y = f(\theta) + \epsilon$$

$$\Rightarrow y \approx f(\theta_0) + \frac{\partial f}{\partial \theta} \Big|_{\theta_0} (\theta - \theta_0) + \epsilon$$

$$\Rightarrow \underbrace{y - f(\theta_0)}_{r_0 \leftarrow "y"} \approx \frac{\partial f}{\partial \theta} \underbrace{(\theta - \theta_0)}_{\delta \leftarrow " \beta " } + \epsilon$$

$$\frac{\partial f}{\partial \theta} = \begin{bmatrix} \frac{\partial f}{\partial \theta_1} \Big|_1 & \frac{\partial f}{\partial \theta_2} \Big|_1 & \dots & \frac{\partial f}{\partial \theta_p} \Big|_1 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial \theta_1} \Big|_n & \frac{\partial f}{\partial \theta_2} \Big|_n & \dots & \frac{\partial f}{\partial \theta_p} \Big|_n \end{bmatrix}$$

↔ "X"

$$\frac{\partial f}{\partial \theta} \equiv J \text{ (Jacobian)}$$

$$\Rightarrow \delta = J^T (J^T J)^{-1} r$$

1. Gissa θ_0
2. Beräkna $f(\theta_0)$, $\frac{\partial f}{\partial \theta} \Big|_{\theta_0}$
3. Beräkna r , J
4. δ_0 kan beräknas
5. $\delta_0 = \theta - \theta_0 \Rightarrow \theta_1 = \delta_0 + \theta_0$

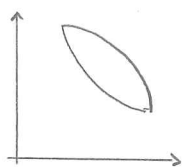
Konfidensintervall

När mätdata jämförs med teori vill vi veta hur säkra parametrarna är. När vi vill veta hur säker en enskild parameter är används ett individuellt konfidensintervall. Vill vi istället jämföra hela modellen och därmed alla parametrar används ett sammansatt konfidensintervall (konfidensyta).

Hur bra prediktion kan vi förvänta oss av modellen?

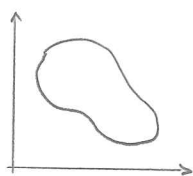
Om vi har ett 95% konfidensintervall finns det sanna värdet inom intervallet 19 gånger av 20 då undersökningen görs.

Icke linjär modell:



Korrekt konfidensgrad

Approximativ form



Approximativ konfidensgrad

Korrekt form