Regression forts.

Ex Sintring av mesa. Modell: $S = S_0 \left(1 - (kt)^{1/\phi}\right)$

Parametrar: ϕ, k, S_0. Värdet på ϕ är en indikator på vilken sintringsmekanism som är mest trolig. Hur säker är skattningen av dessa parametrar? Beräkna konfidensintervall!

1. Bestäm parametrar på lämpligt sätt, genom att minimera SSE

$$\min SSE(\theta) = \min \sum_{n=1}^{N} (y_n - f(x_n, \theta))^2$$

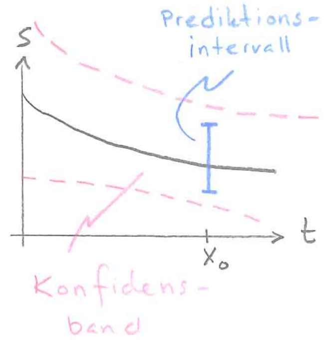
2. Skatta spridningen σ med $s = \sqrt{SSE/(N-p)}$

3. Linjärt ⇒ Behöver X
   Ickelinjärt ⇒ Behöver J

4. Individuellt konfidensintervall: $\theta_p \pm se(\theta_p) \cdot t(N-P, \alpha/2)$
   där $se(\theta_p) = s\sqrt{\left((J^TJ)^{-1}\right)_{pp}}$ är standardfelet.

5. Intervall för prediktion

$$f(x,\theta) \pm s\sqrt{j^T(J^TJ)^{-1}j}\ \sqrt{P \cdot F(P, N-P, \alpha)} \Rightarrow$$

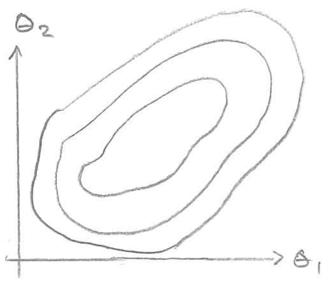


I en punkt bestäms prediktionsintervallet av

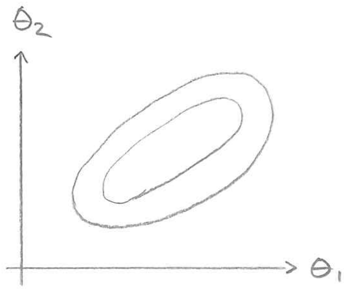
$$f(x_0, \theta) \pm s\sqrt{j_0^T (J^TJ)^{-1} j_0}\ t(N-P, \alpha/2),$$

$$j = \frac{\partial f(x,\theta)}{\partial \theta^T}$$

6. Sammansatt konfidensintervall. Ickelinjära modeller kan ej ha både korrekt konfidensgrad och korrekt form.



Korrekt form
Approximativ konfidensgrad



Korrekt konfidensgrad
Approximativ form

Korrelation mellan parametrar

Uppstår då flera parametrar beskriver samma sak, och beror ofta på hur försöken utförts. För linjära modeller är korrelationen enbart beroende av X. Korrelation ses i konfidensområdets utseende; ett avlångt, utdraget konfidensområde tyder på hög korrelation.

Korrelationen kan även beskrivas av korrelationsmatrisen C:

$$C_{ij} = \frac{(J^TJ)^{-1}_{ij}}{\sqrt{(J^TJ)^{-1}_{ii}\,(J^TJ)^{-1}_{jj}}}$$

$C_{ij} \approx 0 \Rightarrow$ Låg korrelation

$C_{ij} \approx \pm 1 \Rightarrow$ Hög korrelation

Transformering av modeller

En ickelinjär modell, ex $r(x,\theta) = \frac{\theta_1 x}{\theta_2 + x}$, kan ibland transformeras till en linjär modell, ex $y = \frac{1}{r} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1}\cdot\frac{1}{x} = \beta_1 + \beta_2 u$

Dagens frågor

- Hur kontrolleras att förutsättningarna för minsta kvadratmetoden är uppfyllda?
- Hur identifieras vad som är viktigast att förbättra i modellen?
- Hur bestäms när modellen är tillräckligt bra?

Antaganden bakom minsta kvadratmetoden:

- y är modell + störning
- x är utan fel (mäts perfekt)
- Störningarna är normalfördelade
- Störningarna har konstant varians, vilket undersöks genom upprepade försök (spridningsjämförelse) eller residualanalys.

Vilken form är bäst av den transformerbara modellen ovan?

$r(x,\theta) = \frac{\theta_1 x}{\theta_2 + x}$ eller $y = \beta_1 + \beta_2 u$?

– Den som bäst uppfyller antagandena, speciellt om konstant varians, bör väljas!

Residualanalys

För att undersöka en modell kan residualen $e = (y - \hat{y})$ plottas mot

- Tidsordning
- Beräknade värden y
- Alla oberoende variabler x
- Alla tänkbara variabler som ej inkluderas i modellen
- "Ingenting" - Är den normalfördelad?

Utvecklade former av residual analys:

Normalisering: $\frac{e}{s}$, där s är spridningen, gör det lättare med normal fördelnings kontroll och att upptäcka avvikande data

Studentisering: $\frac{e}{\sqrt{(1-H_{ii})s^2}}$, där $H = X(X^TX)^{-1}X^T$, gör kontrollen av konstant varians mer korrekt

Analys av residualplottar

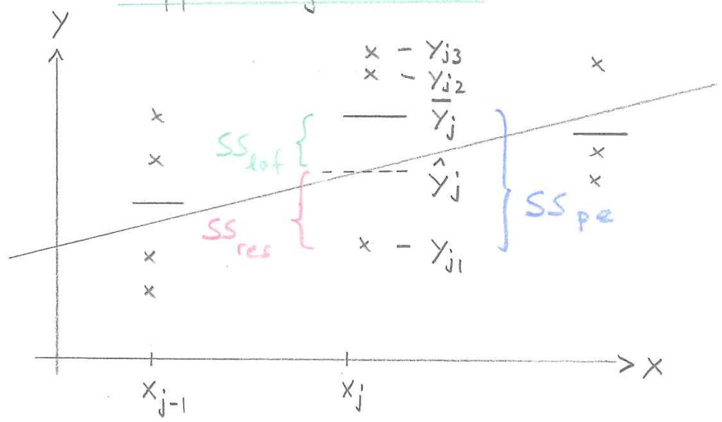
- Icke konstant varians - transformera y, använd viktad residual.
- Linjär trend - linjär term saknas.
- "Krökt" band - kvadratisk term saknas.

"Lack-of-fit" - analys

En undersökningsmetod som undersöker om modellen är signifikant felaktig. Baseras på jämförelse av två sätt att skatta spridning.

Om modellen är väntevärdesriktig är den OK, dvs modellen förutsäger medelvärdet korrekt. Skattning av spridningen är korrekt endast om modellen är väntevärdesriktig. Detta kan kontrolleras genom att skatta σ^2 på ett modelloberoende sätt, vilket kan göras om vi har flera mätningar i varje mätpunkt x_j.

Uppdelning av SS



$$SS_{res} = SS_{pe} + SS_{lof}$$

$$\Rightarrow \underbrace{\sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{ji}-\hat{y}_j)^2}_{SS_{res}} = \underbrace{\sum_{j=1}^{m}\sum_{i=1}^{n_j}(y_{ji}-\bar{y}_j)^2}_{SS_{pe}} + \underbrace{\sum_{j=1}^{m} n_j(\hat{y}_j-\bar{y}_j)^2}_{SS_{lof}}$$

Lack-of-fit - test

Om $\dfrac{SS_{lof}/\nu_{lof}}{SS_{pe}/\nu_{pe}} > F(\nu_{lof}, \nu_{pe}, \alpha)$ har vi

signifikant lack-of-fit, dvs modellen kan förbättras!