Polynomial Regression & Weighted Least Squares

17.1

Should be 17
16 in R

Curvy Trends

- TRANSFORM THE IV

If the scatterplot or residual plot suggests $y = g(x)$ + noise, calculate $g(x_i)$ for $i=1,\dots,n$ & fit

$$Y_i = \beta_0 + \beta_1 g(x_i) + \epsilon_i$$

Show rock example

- TRANSFORM THE DV, but

(−) If several IVs residuals show only curvy trend with only one of them, this can INTRODUCE curvy trends in the others

(−) will affect distribution of Y

Sometimes OKAY Lognormal distribution is positively skewed

log is normal, so if

$$Y_i = \alpha_0 x_i^{\alpha_1} \delta_i \quad , \quad \delta_i \sim \text{lognormal},$$

$$\log Y_i = \log \alpha_0 + \alpha_1 \log x_i + \log \delta_i$$

THIS could fix outliers & unequal variance

- POLYNOMIAL REGRESSION 17.2

Compute additional IVs x^2, x^3 etc

Fit: $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$

If 2 bends, add x^3
3 add x^4 etc

This is a model of ignorance about the form of the curvilinear function $g(x)$. If you knew what it was, you'd just transform the IV that way

Justify in terms of Taylor's Theorem

$$g(x) = g(a) + g'(a)(x-a) + g''(a)\frac{(x-a)^2}{2!} + \dots + R$$

Disregard the remainder & expand the $(x-a)^k$ terms

SHOW CURVY OVERHEADS

- Genuine non-linear regression, like

$$Y_i = \alpha_0 + \alpha_1 \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) + \varepsilon_i$$

Not linear in parameters α_0 α_1 β_0 β_1

Also curvilinear

ROCKS

Unequal variance: Sec 7.8 p. 164 →

17.3

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 V)$$

V pos def, known

TRANSFORM THE DATA

$$V^{-1/2} Y = V^{-1/2} X\beta + V^{-1/2}\varepsilon$$

New error term $V^{-1/2}\varepsilon \sim N(0, \sigma^2 V^{-1/2} V V^{-1/2})$
$= N(0, \sigma^2 I_n)$

And use new DV $V^{-1/2}Y$ & new IVs $V^{-1/2}X$ to estimate β as usual

$$\hat{\beta} = \left((V^{-1/2}X)'V^{-1/2}X\right)^{-1}(V^{-1/2}X)'V^{-1/2}Y$$

$$= (X'V^{-1}X)^{-1}X'V^{-1}Y \qquad (7.63)$$

p. 165

etc.

17.4

Non-diagonal V may be useful in time series analysis, but most commonly, ε_i are still independent, but with $V(\varepsilon_i) = \frac{\sigma^2}{w_i}$. Then

Variance of SAMPLE MEAN

$$V = \begin{pmatrix} 1/w_1 & & & 0 \\ & 1/w_2 & & \\ & & \ddots & \\ 0 & & & 1/w_n \end{pmatrix} \text{ \& } V^{-1/2} = \begin{pmatrix} \sqrt{w_1} & & & 0 \\ & \sqrt{w_2} & & \\ & & \ddots & \\ 0 & & & \sqrt{w_n} \end{pmatrix}$$

$$V^{-1/2} Y = \begin{pmatrix} \sqrt{w_1}\, Y_1 \\ \sqrt{w_2}\, Y_2 \\ \vdots \\ \sqrt{w_n}\, Y_n \end{pmatrix}$$ so that $\mathrm{var}(\sqrt{w_i}\, Y_i) = w_i \frac{\sigma^2}{w_i} = \sigma^2$

$V^{-1/2} X$ multiplies row i of X by $\sqrt{w_i}$ also.

& the least squares problem is to minimize

$$Q = \sum_{i=1}^{n} \left(\sqrt{w_i}\, Y_i - \sqrt{w_i}\, \beta_0 - \beta_1 \sqrt{w_i}\, x_{i1} - \cdots - \beta_p \sqrt{w_i}\, x_{ip}\right)^2$$

$$= \sum_{i=1}^{n} w_i \left(Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}\right)^2$$

"weights" are inversely proportional to $V(Y_i)$

THE SMALLER THE VARIANCE, THE MORE IT COUNTS.

Show can scale fix w/