Statistics 203: Introduction to Regression and Analysis of Variance

Nonlinear regression

Jonathan Taylor

Today's class

- Nonlinear regression.
- Examples in R.

Nonlinear regression models

- We have usually assumed regression is of the form

$$Y_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ij} + \varepsilon_i.$$

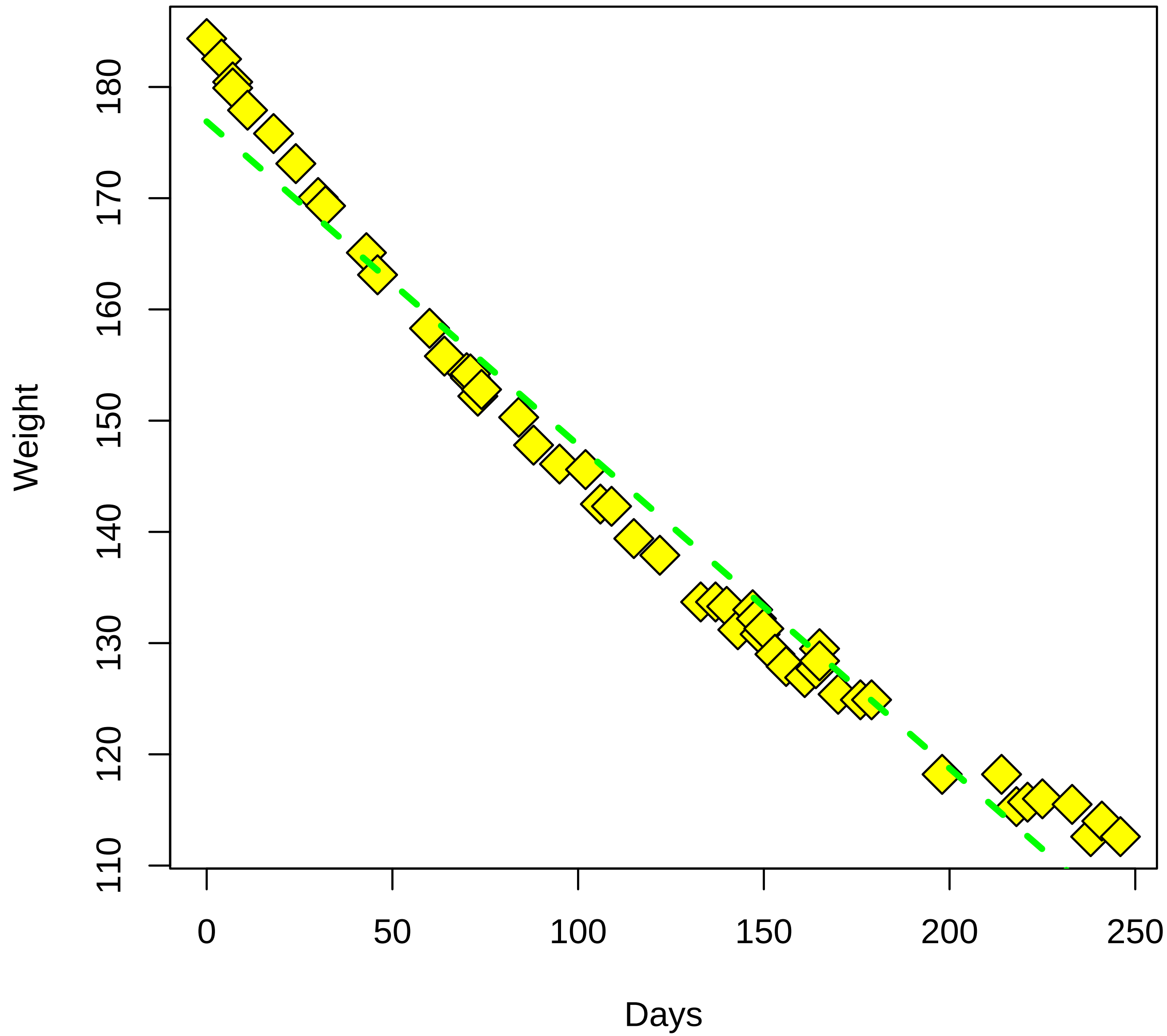
- Or, the regression function

$$f(x, \beta) = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_j$$

  is *linear* in beta.
- Many real-life phenomena can be parameterized by *non-linear* regression functions. Example:
  - Radioactive decay: half-life is a non-linear parameter

$$f(t, \theta) = C \cdot 2^{-t/\theta}.$$

Weight loss data



What to do?

- In some cases, it is possible to linearize to get the original model.
- Suppose
$$Y_i = C \cdot 2^{-X_i/\theta} \varepsilon_i,$$
then
$$\log(Y_i) = C' - X_i/\theta + \varepsilon_i^*.$$
If ε_i^* have approximately the same variance, than it looks like original model.
- What if
$$Y_i = C \cdot 2^{-X_i/\theta} + \varepsilon_i$$
where the ε_i's have the same variance?

Delta method

- The delta method tells us that

$$\text{Var}(f(Y_i)) \simeq f'(\mathbb{E}(Y_i))^2 \text{Var}(Y_i).$$

- In this case

$$\text{Var}(\log(Y_i)) \simeq \frac{1}{C^2 2^{-2X_i/\theta}} \sigma^2.$$

- Could use weighted least-squares if θ was known.
- One possibility: IRLS.
  - Starting at some initial $\widehat{\theta}^j, j \geq 0$, regress $\log(Y_i)$ onto X_i solve for $\widehat{\beta}_1^{j+1}$, and set

$$\widehat{\theta}_0^{j+1} = -\frac{1}{\widehat{\beta}_1^{j+1}}$$

Nonlinear regression

- Alternatively,

$$(\widehat{C}, \widehat{\theta}) = \underset{(C,\theta)}{\text{argmin}} \sum_{i=1}^{n} (Y_i - C2^{-X_i/\theta})^2.$$

- In general, given a possibly non-linear regression function $f(x, \theta), \theta \in \mathbb{R}^p$: we can try to solve

$$\widehat{\theta} = \underset{\theta}{\text{argmin}} \sum_{i=1}^{n} (Y_i - f(x_i, \theta))^2$$

  where x_i is the i-th row of the design matrix.
- Techniques: usually use a steepest descent approach instead of Newton-Raphson.

Nonlinear regression: details

- Given an initial value $\widehat{\theta}^0$

$$f(x_i,\theta) \simeq f(x_i,\widehat{\theta}^0) + \sum_{j=1}^{p} \frac{\partial f}{\partial \theta_j}\Big|_{(x_i,\theta)=(x_i,\widehat{\theta}^0)} (\theta_i - \theta_i^0) = \eta_i^0(\theta).$$

- In matrix form

$$\eta^0(\theta) = \omega^0 + Z^0\theta$$

where

$$Z_{ij}^0 = \frac{\partial f}{\partial \theta_j}\Big|_{(x,\theta)=(x_i,\widehat{\theta}^0)} = \frac{\partial \eta_i}{\partial \theta_j}\Big|_{\theta=\widehat{\theta}^0}$$

$$\omega_i^0 = f(x_i,\theta_0) - \sum_{j=1}^{p} \theta_i^0 Z_{ij}^0$$

- The vector $\eta^0(\theta)$ is our approximation to the regression surface, we want to choose θ to get as close to Y as possible. Project onto “tangent space.”

Iteration & Distribution

- Want to choose $\widehat{\theta}^1$ as follows

$$\widehat{\theta}^1 = \underset{\theta}{\text{argmin}} \sum_{i=1}^n (Y_i - \omega_i^0 - Z^0\theta)^2$$

- Solution

$$\widehat{\theta}^1 = \left(Z^{0\,t} Z^0\right)^{-1} Z^{0\,t}(Y - \omega^0).$$

- Given $\widehat{\theta}^j$, set

$$\widehat{\theta}^{j+1} = \left(Z^{j\,t} Z^j\right)^{-1} Z^{j\,t}(Y - \omega^j).$$

- Repeat until convergence.
- Approximate distribution of $\widehat{\theta}$:

$$\widehat{\theta} \sim N(\theta, \widehat{\sigma}^2(\widehat{Z}^t\widehat{Z})^{-1}),$$

where

$$\widehat{\sigma}^2 = \sum_{i=1}^n (Y_i - f(x_i, \widehat{\theta}))^2/(n-p).$$

Confidence intervals

- If θ is restricted, say $\theta \geq 0$ the asymptotic confidence intervals may be inaccurate (may overlap into negative numbers).
- `library(MASS)` provides another method to obtain confidence intervals based on "inverting" an F-test.
- Basic idea, confidence interval for θ_1: for each fixed value $\theta_{1,0}$ we could compute the "extra sum of squares" between the unrestricted model and the model with θ_1 fixed at $\theta_{1,0}$.

$$F(\theta_{1,0}) = \frac{SSE_{\theta_1=\theta_{1,0}}(\widehat{\theta}_{2:p}) - SSE(\widehat{\theta})}{\widehat{\sigma}^2} \sim F_{1,n-p}$$

  at least approximately under $H_0 : \theta_1 = \theta_{1,0}$.
- Or,

$$T(\theta_{1,0}) = \mathrm{sign}(\widehat{\theta} - \widehat{\theta}_{1,0})\sqrt{F(\theta_{1,0})} \sim t_{n-p}.$$

- Confidence interval:

$$\{\theta_1 : -t_{n-p,1-\alpha/2} < T(\theta_1) < t_{n-p,1-\alpha/2}\}.$$

Weight loss data