

## Scatterplot Smoothing

- Scatterplot smoothing is an Exploratory Data Analysis method
- Multiple versions are available
  - Robust methods (`lowess`–locally weighted scatterplot smoothing)
  - Non-robust methods (`loess`–local regression–and `scatter.smooth`)
  - `loess` is the most flexible; `scatter.smooth` is used primarily as a graphing function

## Scatterplot Smoothing

### *Weighted Least Squares Regression*

- Simple linear regression estimates the slope and intercept of the line  $y = \beta_0 + \beta_1 x$  by minimizing the least squares function:

$$Q = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$$

- Weighted least squares regression minimizes the *weighted* least squares function:

$$Q_w = \sum_{i=1}^n w_i [y_i - (\beta_0 + \beta_1 x_i)]^2$$

- WLS is traditionally used to downweight observations with more variability

## Scatterplot Smoothing

### *Local Regression*

- Local regression generates a separate (local) regression line for each  $x_i$  using only *part* of the sample. These  $n$  subsets are centered at each  $x_i$  in turn. Typical subsets comprise 1/3 to 2/3 of the data centered at each  $x_i$  in turn.
- In local regression, we are no longer focussed on each line, only the predicted value of each line at a given  $x_i$ .
- After computing predicted values from all these separate subset regressions, we connect these values to create a *smoothed curve*.

## Scatterplot Smoothing

### *Weight functions for Local Regression*

- We use a different prediction equation at each  $x_i$
- These prediction equations give more weight to observations *near*  $x_i$
- `loess` uses the *tricubic function* to weight each observation in the subset  $S_i$ :

$$w(x) = (1 - \|x - x_i\|/\max_{S_i}\|x - x_i\|)^3, \quad \|x - x_i\| < \max_{S_i}\|x - x_i\|$$

## Scatterplot Smoothing

`loess` in R

- Important arguments include `formula`, `span`, `degree`
- Model objects such as these have a lot of attributes
- Inference is possible, though simple concepts in linear regression, such as model `df`, are now more subtle
- Use `predict` to overlay a smoothed line

## Scatterplot Smoothing

### *Conclusions*

- `scatter.smooth` can be used as a shortcut for plotting
- `lowess` is iterative and has fewer options
  - WLS local regression
  - rescale weights using residuals (observations with large residuals are downweighted)
  - New WLS local regression using rescaled weights
  - etc