Linear Regression Models
P8111

Lecture 13

Jeff Goldsmith
March 3, 2016

THE DEPARTMENT OF BIOSTATISTICS
Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Today's Lecture

- Model selection vs. model checking
- Continue with model checking (regression diagnostics)

Model selection vs. model checking

In a model of the form

$$y|\boldsymbol{x} = f(\boldsymbol{x}) + \epsilon$$

model selection focuses on how you construct $f(\cdot)$; model checking asks whether the ϵ match the assumed form.

Model checking

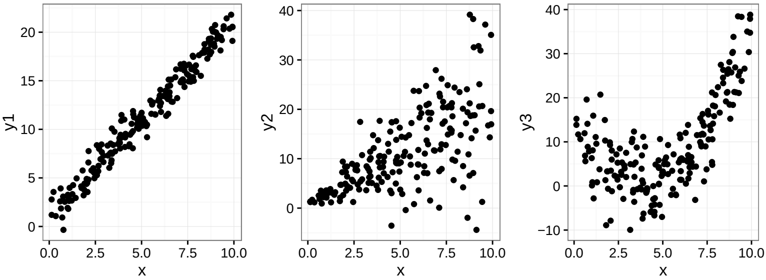
Two major areas of concern:

- Global lack of fit, or general breakdown of model assumptions
  - Linearity
  - Unbiased, uncorrelated errors $E(\epsilon|x) = E(\epsilon) = 0$
  - Constant variance $Var(y|x) = Var(\epsilon|x) = \sigma^2$
  - Independent errors
  - Normality of errors
- Effect of influential points and outliers

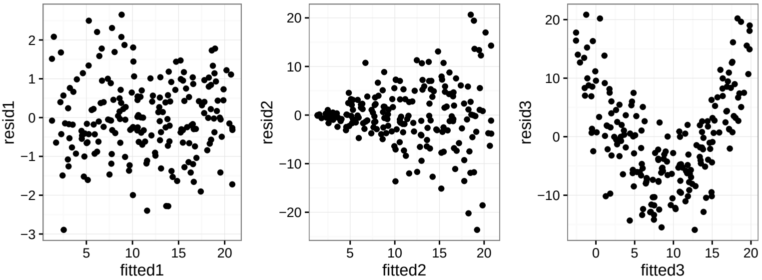
Model checking

- Global lack of fit, or general breakdown of model assumptions
  - Residual analysis – QQ plots, residual plots against fitted values and predictors
  - Adjusted variable plots
- Effect of influential points and outliers
  - Measure of leverage, influence, outlying-ness

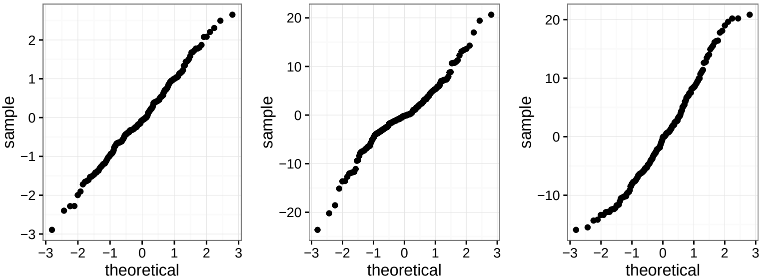
Some data plots



Some residual plots



Checking Normality assumption



Non-constant variance

What to do ...

- Nothing; just use least squares and bootstrap
- Use weighted LS, GLS (later)
- Use a variance stabilizing transformation

Variance-stabilizing transformation

Suppose y is strictly positive, $\mu = E(y|x)$, $Var(y|x) = \sigma^2 g(\mu)$

- Replace y with $y^* = T(y)$ such that $Var(y^*|x)$ is approximately constant
- Delta method says $Var(T(y)) = (T'(\mu))^2 \sigma^2 g(\mu)$

Variance-stabilizing transformation

To get constant variance, we want

$$\begin{aligned}
(T'(\mu))^2 g(\mu) &= k^2 \quad \text{(constant)} \\
\Rightarrow T'(\mu) &= \frac{k}{\sqrt{g(\mu)}} \\
\Rightarrow T(\mu) &= \int \frac{k}{\sqrt{g(\mu)}} d\mu \\
\Rightarrow T(y) &= \int \frac{k}{\sqrt{g(y)}} dy
\end{aligned}$$

So the transformation necessary to stabilize the variance really depends on the variance function itself, e.g. $g(\cdot)$

Variance-stabilizing transformation examples

- Example 1: If $Var(y|x) = \sigma^2\mu^2$, i.e. if $g(y) = y^2$, $T(y) =?$

- Example 2: If $Var(y|x) = \sigma^2\mu$, i.e. if $g(y) = y$, $T(y) =?$

Isolated points

Points can be isolated in three ways

- Leverage point – outlier in x
- Outlier – outlier in $y|\boldsymbol{x}$
- Influential point – a point that largely affects $\boldsymbol{\beta}$
  - Deletion influence; $|\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}|$
  - Basically, a high-leverage outlier

Leverage is measured by the hat matrix, outlying-ness by the residual

Quantifying leverage

We measure leverage (the "distance" of $\boldsymbol{x}_i$ from the distribution of $\boldsymbol{x}$) using

$$h_{ii} = \boldsymbol{x}_i^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_i$$

where h_{ii} is the $(i,i)^{th}$ entry of the hat matrix.

Leverage

Some notes about the hat matrix

- $\sum_i h_{ii} \stackrel{def}{=} tr(\boldsymbol{H}) = (p+1)$

(Note – the trace of the hat matrix generalizes to non-parametric methods, where you don't have a specific number of parameters to count. This is a useful measure of "model size" or "effective degrees of freedom" in these cases.)

Leverage

Some notes about the hat matrix

- $\hat{y}_i = \sum_j h_{ij} y_j$

- $\sum_i h_{ij} = \sum_j h_{ij} = 1$

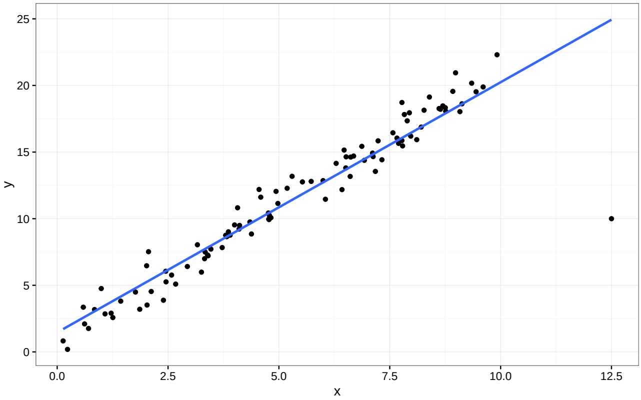
These mean that h_{ii} is the weight given to y_i in determining $\hat{y}_i$

Leverage

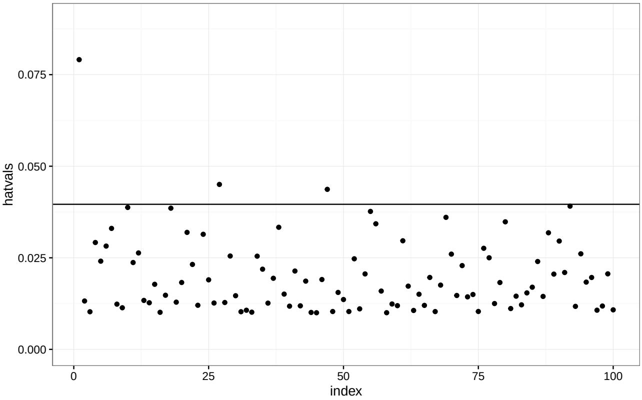
What counts as "big" leverage?

- Average leverage is $(p+1)/n$
- Typical rules of thumb are $2(p+1)/n$ or $3(p+1)/n$
- Leverage plots can be useful as well

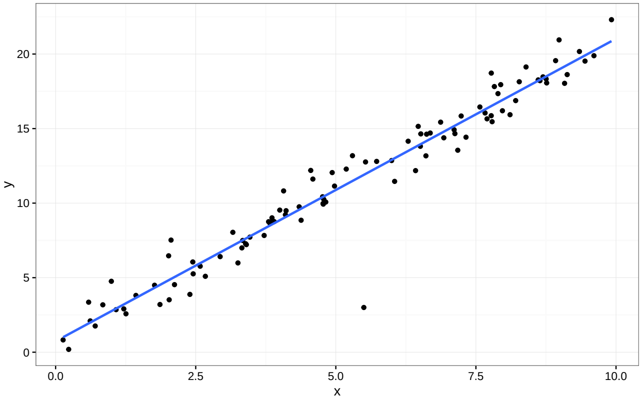
Leverage plot



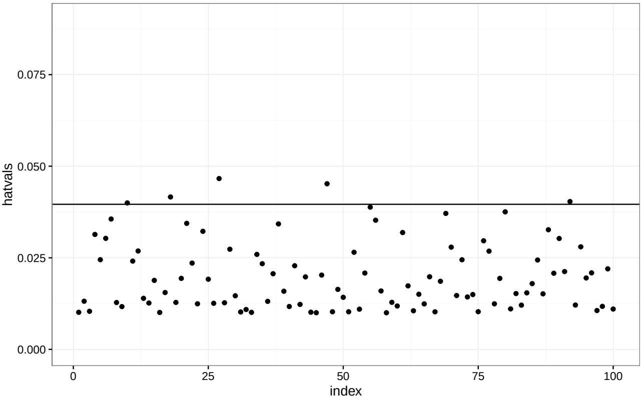
Leverage plot



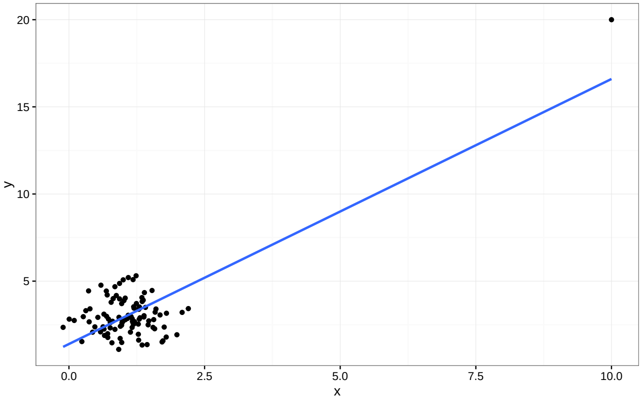
Leverage plot



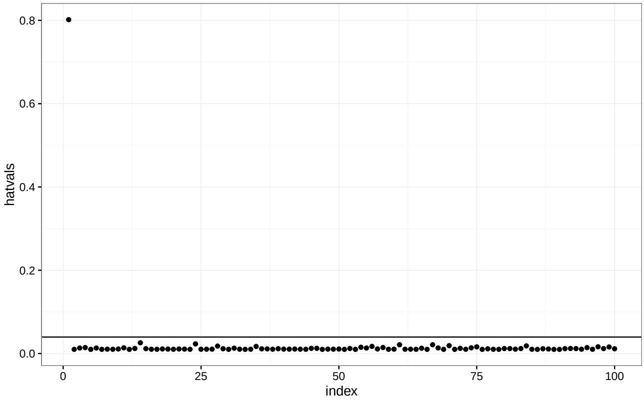
Leverage plot



Leverage plot



Leverage plot



Outliers

- When we refer to “outliers” we typically mean “points that don't have the same mean structure as the rest of the data”
- Residuals give an idea of “outlying-ness”, but we need to standardize somehow
- Remember (from last lecture) $Var(\hat{\epsilon}_i) = \sigma^2(1 - h_{ii})$ ...

Outliers

The *standardized* residual is given by

$$\hat{\epsilon}_i^* = \frac{\hat{\epsilon}_i}{\sqrt{Var(\hat{\epsilon}_i)}} = \frac{\hat{\epsilon}_i}{\hat{\sigma}\sqrt{(1-h_{ii})}}$$

The *Studentized* residual is given by

$$t_i = \frac{\hat{\epsilon}_i}{\hat{\sigma}_{(-i)}\sqrt{(1-h_{ii})}} = \hat{\epsilon}_i^* \left(\frac{n-(p+1)}{n-(p+1)-\hat{\epsilon}_i^{*2}}\right)^{1/2}$$

Studentized residuals follow a $t_{n-(p+1)-1}$ distribution.

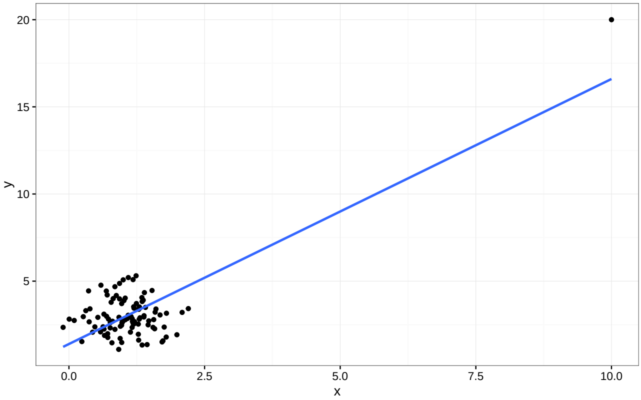
Influence

Specifically, deletion influence

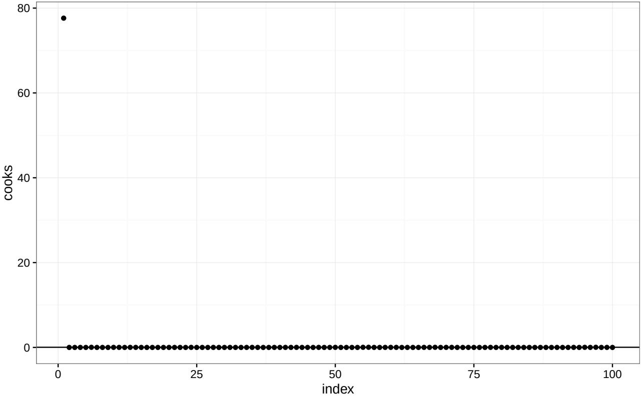
- $|\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}|$
- Cook's distance is

$$
\begin{aligned}
D_i &= \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)})^T (\boldsymbol{X}^T \boldsymbol{X}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)})}{(p+1)\hat{\sigma}^2} \\
&= \frac{(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{(-i)})^T (\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{(-i)})}{(p+1)\hat{\sigma}^2} \\
&= \frac{1}{p+1} \hat{\epsilon}_i^2 \frac{h_{ii}}{1 - h_{ii}}
\end{aligned}
$$

Cook's distance plot



Cook's distance plot



Handy `R` functions

Suppose you fit a linear model in `R`;

- `hatvalues` gives the diagonal elements of the hat matrix h_{ii} (leverages)
- `rstandard` gives the standardized residuals
- `rstudent` gives the studentized residuals
- `cooks.distance` gives the Cook's distances

Today's big ideas

- Model checking

- Suggested reading: Faraway Ch 7