

# ML Week

## 0x02 Linear Regression

Jeff Abrahamson

2–5 novembre 2015

# Linear models

**Problem:**  $\{(x_i, y_i)\}$ .

Given  $x$ , predict  $\hat{y}$ .

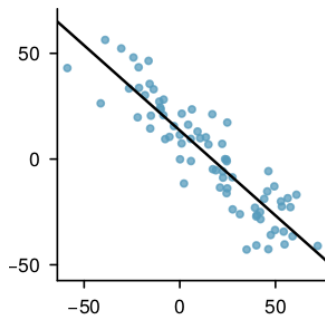
# Linear models

$x$ : **explanatory** or **predictor** variable.

$y$ : **response** variable.

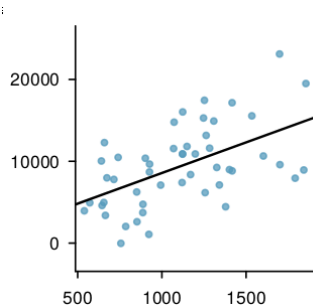
# Linear models

Example:



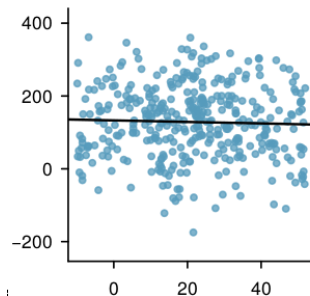
# Linear models

Example:



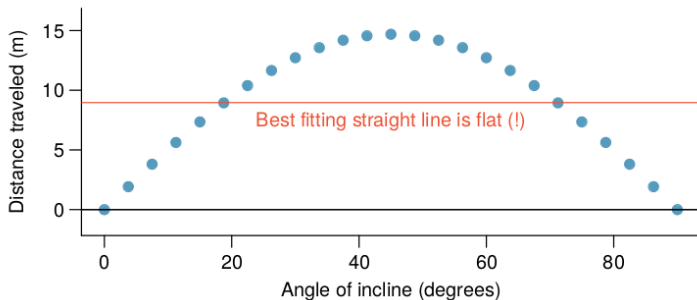
# Linear models

Example:



# Linear models

Example:



# Residuals

What's left over.

$$\text{data} = \text{fit} + \text{residual}$$



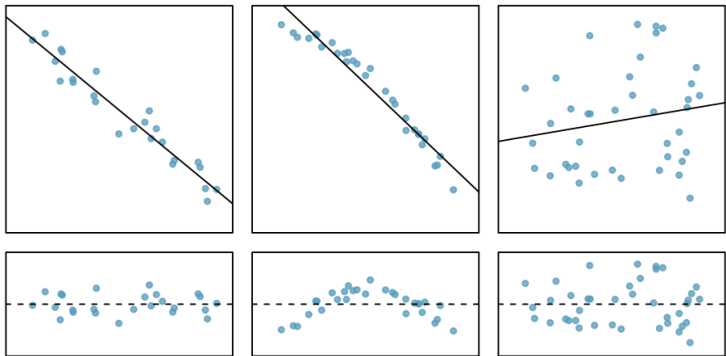
# Residuals

What's left over.

$$y_i = \hat{y}_i + e_i$$

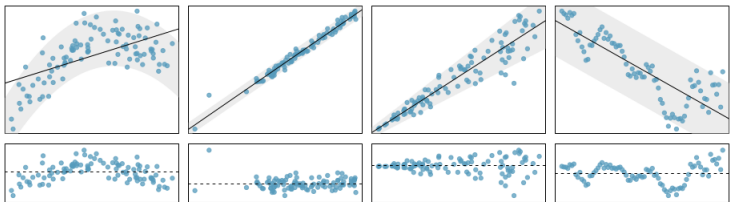
# Residuals

What's left over.



# Residuals

What's left over.



# Residuals

What's left over.

Goal: small residuals.

$$\sum |e_i|$$

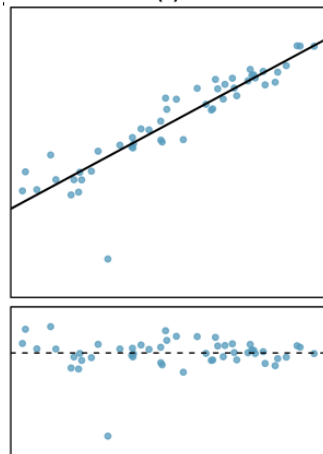
# Residuals

What's left over.

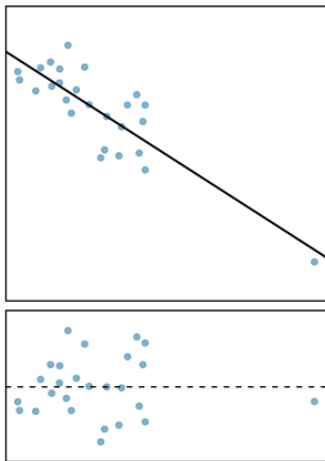
Goal: small residuals.

$$\sum e_i^2$$

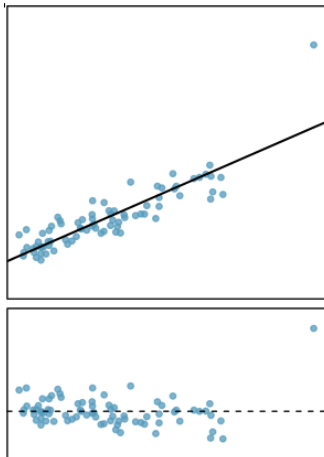
# Outliers



# Outliers

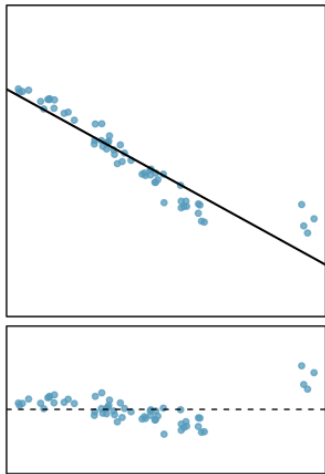


# Outliers

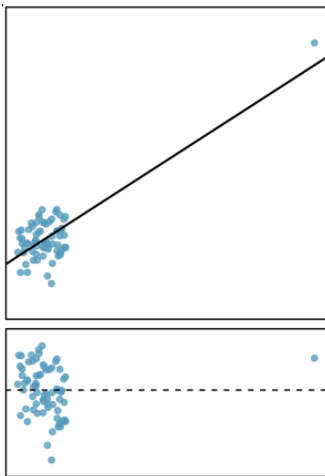




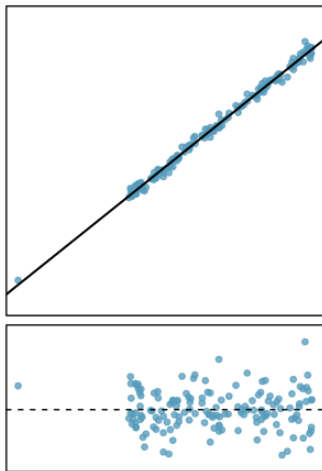
# Outliers



# Outliers

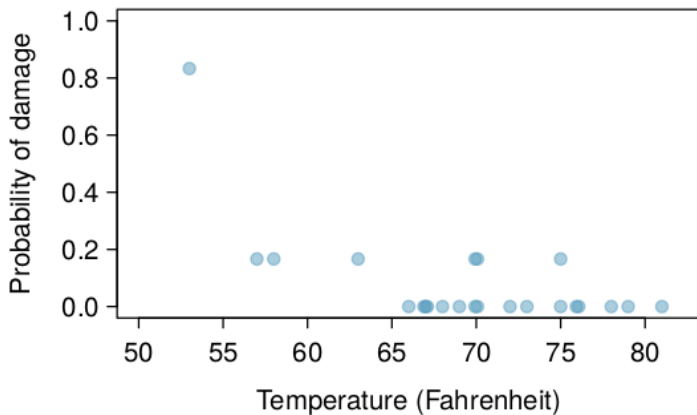


# Outliers

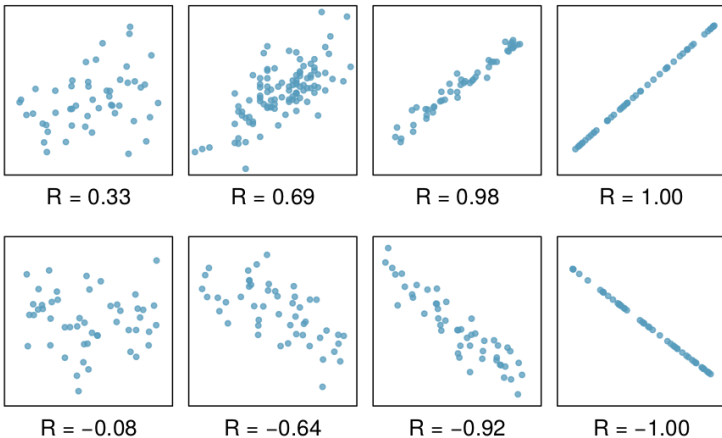


Don't ignore outliers.

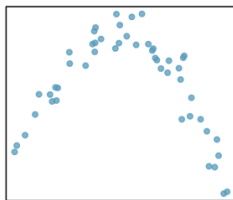
# Outliers



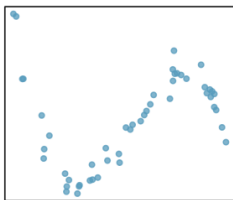
# Correlation



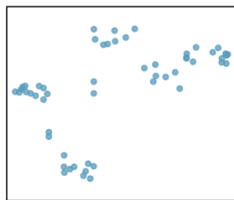
# Correlation



$R = -0.23$



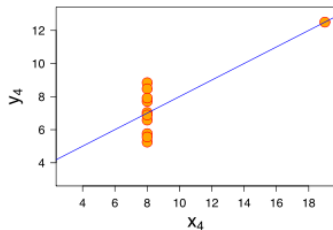
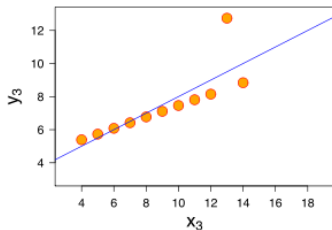
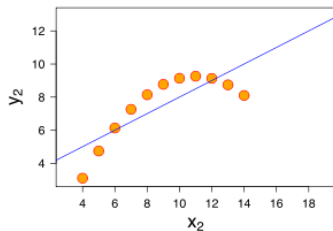
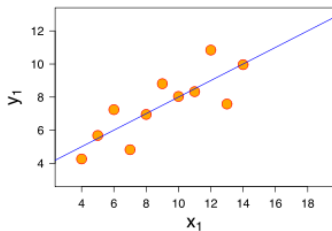
$R = 0.31$



$R = 0.50$

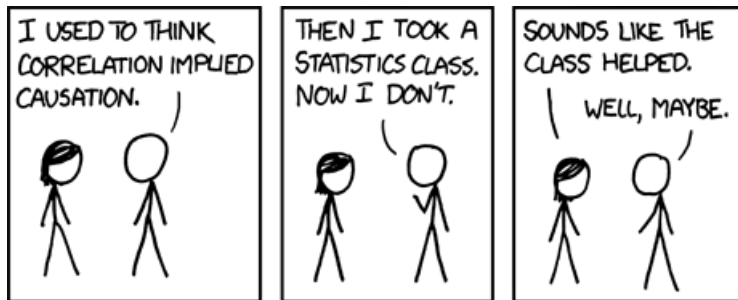
# Correlation

## Anscombe's Quartet





# Correlation does not imply causation



<https://xkcd.com/552/>

# Hypothesis (model)

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

# Cost function

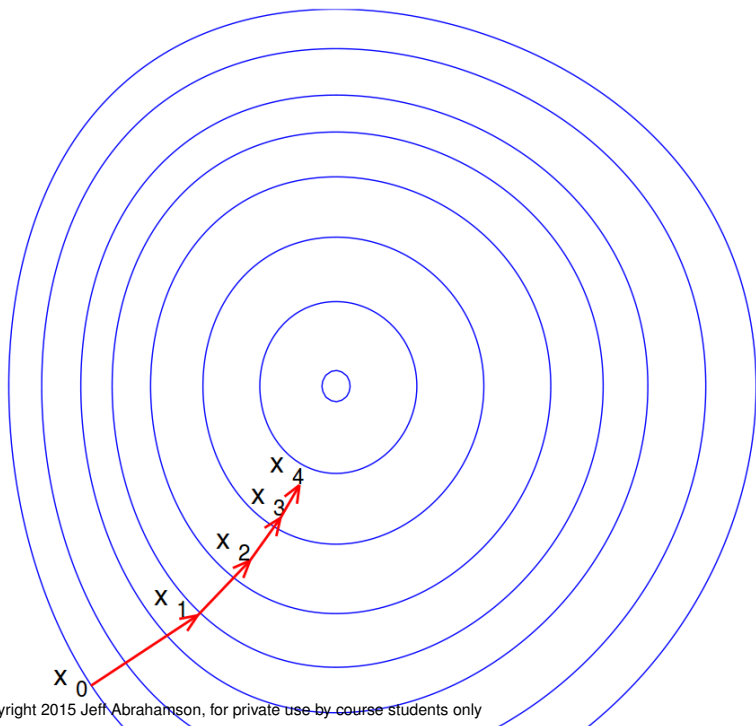
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x_i) - y_i)^2$$

# Gradient descent

$$\begin{cases} \theta_0 & \leftarrow \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\ \theta_1 & \leftarrow \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \end{cases}$$

# Gradient descent

$$\begin{cases} \theta_0 \leftarrow \theta_0 - \frac{\alpha}{m} \sum_{i=1}^m (h_{\theta}(x_i) - y_i) \\ \theta_1 \leftarrow \theta_1 - \frac{\alpha}{m} \sum_{i=1}^m (h_{\theta}(x_i) - y_i) \end{cases}$$



# Hypothesis again

$$\begin{aligned}h_{\theta}(x) &= \theta_0 + \theta_1 x_1 \\&= \theta_0 + \sum_{i=1}^1 \theta_i x_i \\&= [\theta_0, \theta_1] \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \\&= \theta^T x\end{aligned}$$

# Hypothesis (multiple regression)

$$\begin{aligned}h_{\theta}(x) &= \theta_0 + \sum_{i=1}^n \theta_i x_i \\&= [\theta_0, \dots, \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \\&= \theta^T x\end{aligned}$$



# Hypothesis (multiple regression)

$$\begin{aligned}h_{\theta}(x) &= \theta^T x \\ &= \theta^T x^{(1)}\end{aligned}$$

# Hypothesis (multiple regression)

$$X = \begin{pmatrix} \begin{array}{c|c} x^{(1)} & \end{array} & \begin{array}{c|c} x^{(2)} & \end{array} & \cdots & \begin{array}{c|c} x^{(m)} & \end{array} \end{pmatrix} = \begin{pmatrix} x_0^{(1)} & x_0^{(2)} & \cdots & x_0^{(m)} \\ x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(m)} \end{pmatrix}$$

# Hypothesis (multiple regression)

$$\begin{aligned}h_{\theta}(X) &= \theta^T X \\&= [h_0(x^{(1)}), h_0(x^{(2)}), \dots, h_0(x^{(m)})] \\&= \theta^T X\end{aligned}$$

# Hypothesis (multiple regression)

or  $X\theta$  if row vectors. . .

# Cost function (multiple regression)

$$\begin{aligned} J(\theta) &= \frac{1}{2m} \sum_{i=1}^m \left( h_{\theta}(x^{(i)}) - y^{(i)} \right)^2 \\ &= \frac{1}{2m} (X\theta - Y)^T (X\theta - Y) \end{aligned}$$

# Gradient descent (multiple regression)

$$\theta_j \leftarrow \theta_j - \frac{\alpha}{m} \sum_{i=1}^m \left( h_{\theta}(x^{(i)}) - y^{(i)} \right) \cdot x_j^{(i)}$$

for  $j = 1, \dots, n$

# Gradient descent (multiple regression)

$$\theta \leftarrow \theta - \nabla J(\theta)$$

$$\text{where } \nabla = \begin{bmatrix} \frac{\partial}{\partial \theta_0} \\ \frac{\partial}{\partial \theta_1} \\ \vdots \\ \frac{\partial}{\partial \theta_n} \end{bmatrix}$$

# Questions?

`purple.com/talk-feedback`