ML Week

Linear Regression

Jeff Abrahamson

23–24 novembre 2016

Copyright 2016 Jeff Abrahamson, for private use by course students only

Linear models

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

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Linear models

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

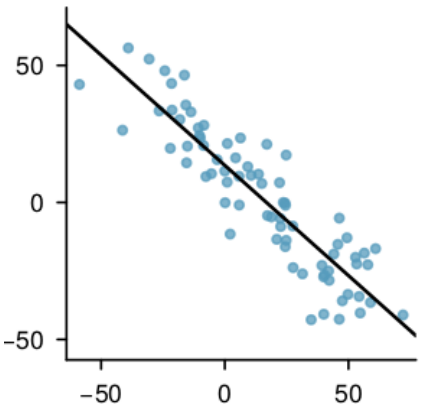
y: **response** variable.

For some reason, we believe a linear model is a good idea.

Copyright 2016 Jeff Abrahamson, for private use by course students only

Linear models

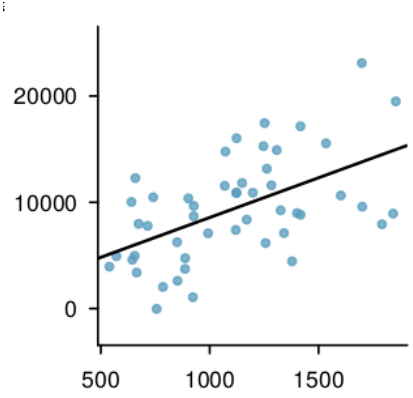
Example:



Copyright 2016 Jeff Abrahamson, for private use by course students only

Linear models

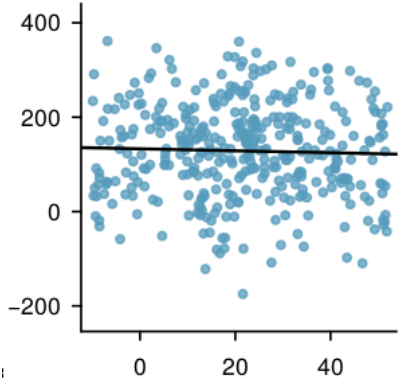
Example:



Copyright 2016 Jeff Abrahamson, for private use by course students only

Linear models

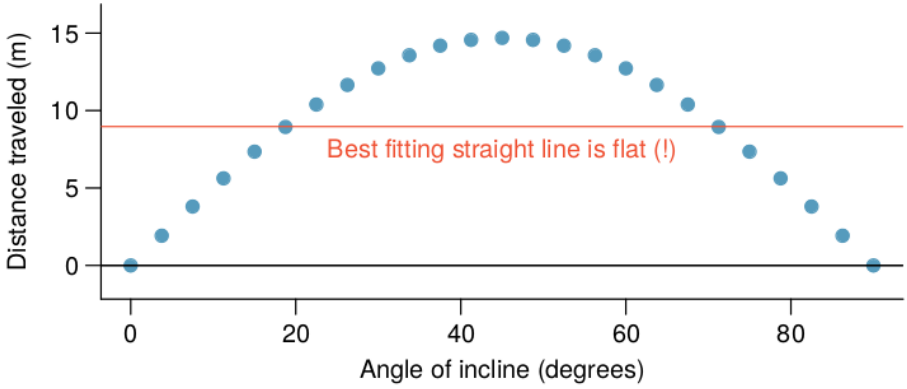
Example:



Copyright 2016 Jeff Abrahamson, for private use by course students only

Linear models

Example:



Copyright 2016 Jeff Abrahamson, for private use by course students only

Residuals

What's left over.

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Residuals

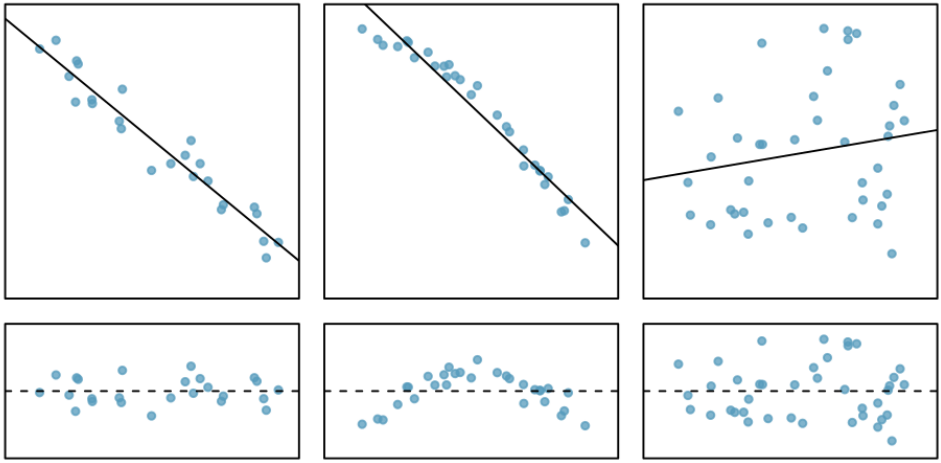
What's left over.

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Residuals

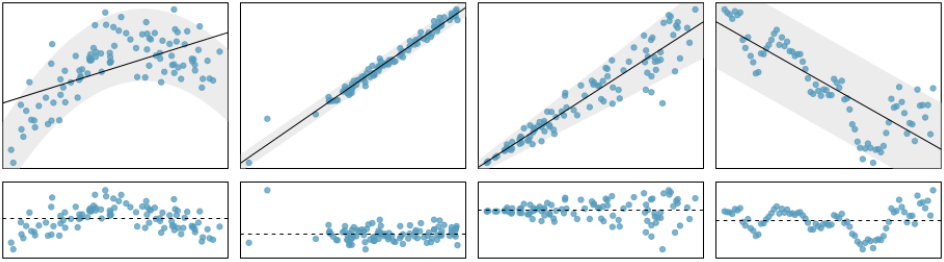
What's left over.



Copyright 2016 Jeff Abrahamson, for private use by course students only

Residuals

What's left over.



Copyright 2016 Jeff Abrahamson, for private use by course students only

Residuals

What's left over.

Goal: small residuals.

$$\sum | e_i |$$

Copyright 2016 Jeff Abrahamson, for private use by course students only

Residuals

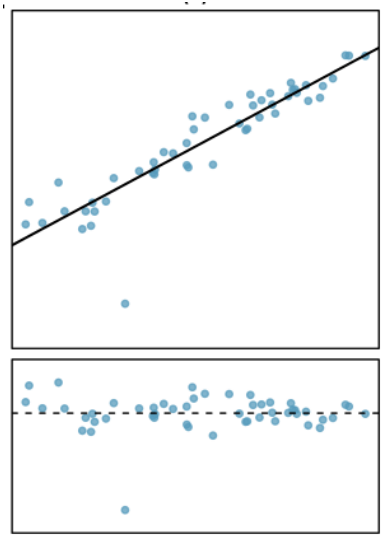
What's left over.

Goal: small residuals.

$$\sum e_i^2$$

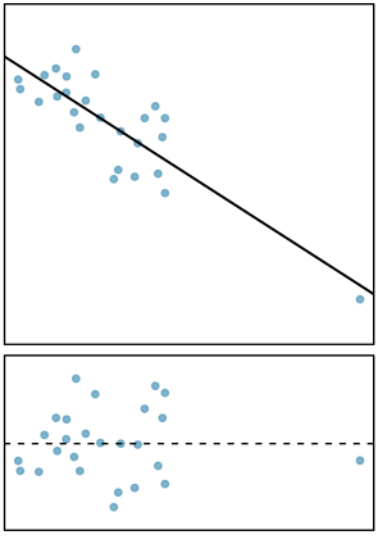
Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers



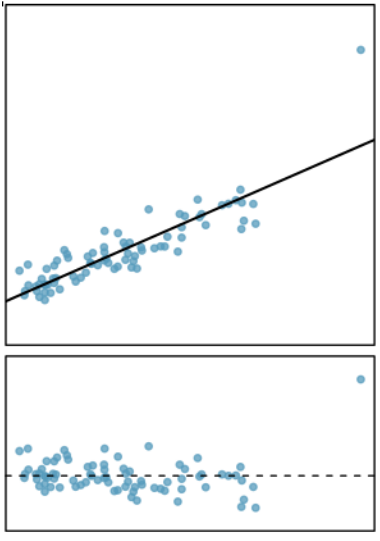
Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers



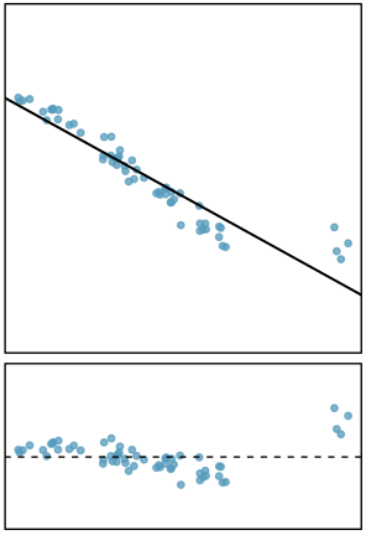
Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers



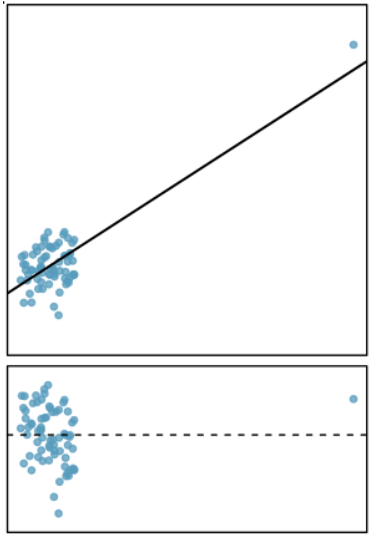
Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers



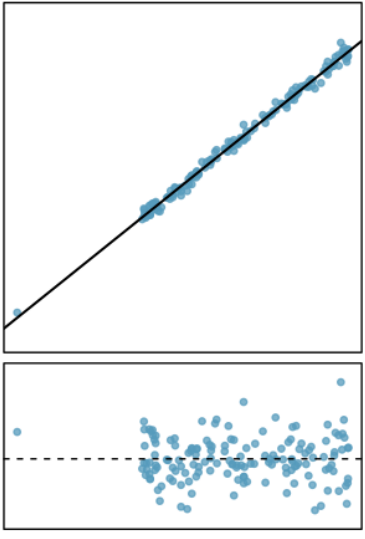
Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers



Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers



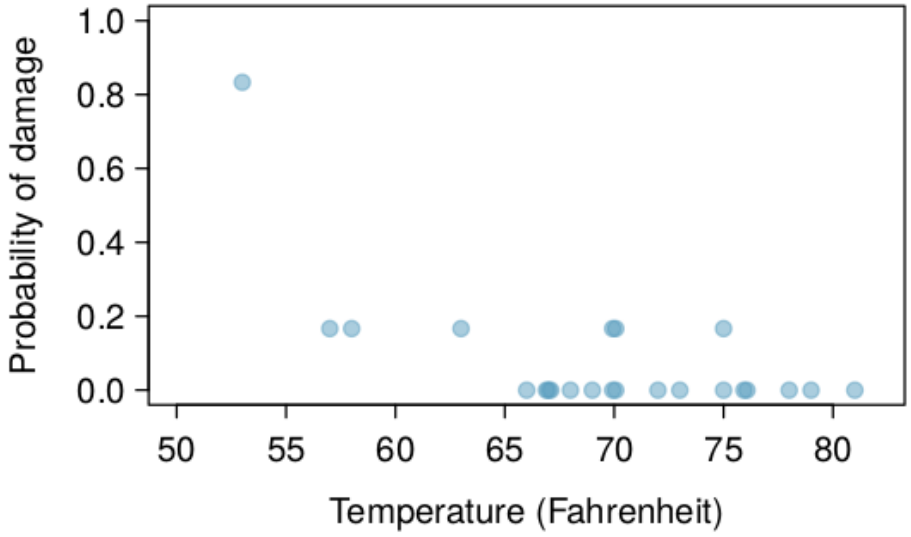
Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers

Don’t ignore outliers.

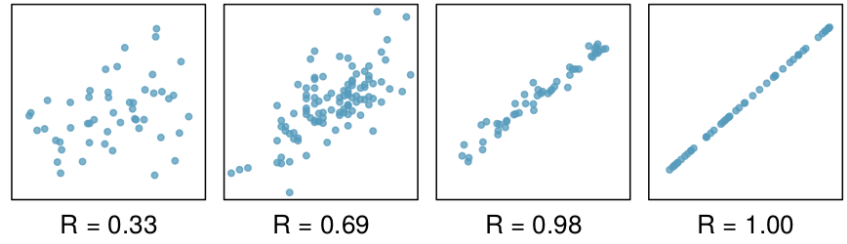
Copyright 2016 Jeff Abrahamson, for private use by course students only

Outliers



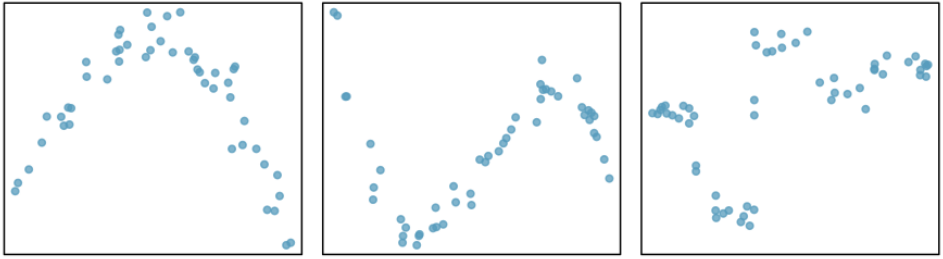
Copyright 2016 Jeff Abrahamson, for private use by course students only

Correlation



Copyright 2016 Jeff Abrahamson, for private use by course students only

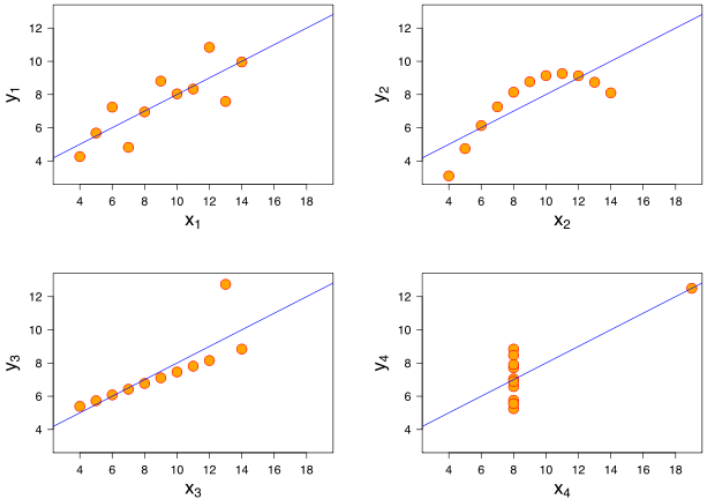
Correlation



Copyright 2016 Jeff Abrahamson, for private use by course students only

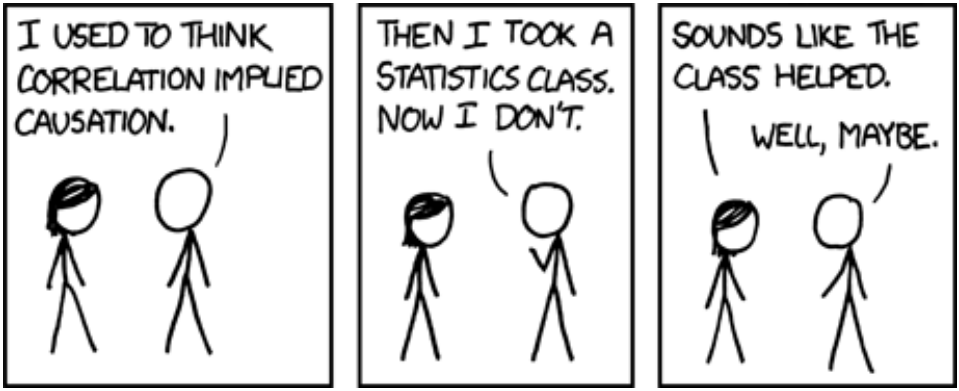
Correlation

Anscombe's Quartet



Copyright 2016 Jeff Abrahamson, for private use by course students only

Correlation does not imply causation



https://xkcd.com/552/

Copyright 2016 Jeff Abrahamson, for private use by course students only

Hypothesis (model)

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Cost function

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Gradient descent

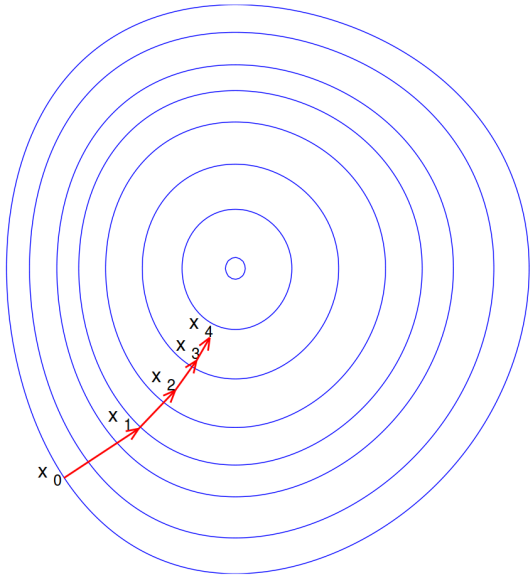
$$\begin{cases} \theta_0 & \leftarrow \theta_0 - \alpha \dfrac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\ \theta_1 & \leftarrow \theta_1 - \alpha \dfrac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \end{cases}$$

Copyright 2016 Jeff Abrahamson, for private use by course students only

Gradient descent

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

Copyright 2016 Jeff Abrahamson, for private use by course students only



Copyright 2016 Jeff Abrahamson, for private use by course students only

Hypothesis again

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Hypothesis (multiple regression)

$$h_\theta(x) = \theta_0 + \sum_{i=1}^{n} \theta_i x_i$$

$$= [\theta_0, \cdots, \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \theta^T x$$

Copyright 2016 Jeff Abrahamson, for private use by course students only

Hypothesis (multiple regression)

$$h_\theta(x) = \theta^T x$$
$$= \theta^T x^{(1)}$$

Copyright 2016 Jeff Abrahamson, for private use by course students only

Hypothesis (multiple regression)

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Hypothesis (multiple regression)

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Hypothesis (multiple regression)

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Cost function (multiple regression)

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

$$= \frac{1}{2m} (X\theta - Y)^T (X\theta - Y)$$

Copyright 2016 Jeff Abrahamson, for private use by course students only

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 $j = 1, \cdots, n$

Copyright 2016 Jeff Abrahamson, for private use by course students only

Gradient descent (multiple regression)

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

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

Copyright 2016 Jeff Abrahamson, for private use by course students only

Questions?

purple.com/talk-feedback

Copyright 2016 Jeff Abrahamson, for private use by course students only