Simple Linear Regression and the Method of Least Squares

Author: Nicholas G Reich, Jeff Goldsmith

This material is part of the ***statsTeachR*** *project*

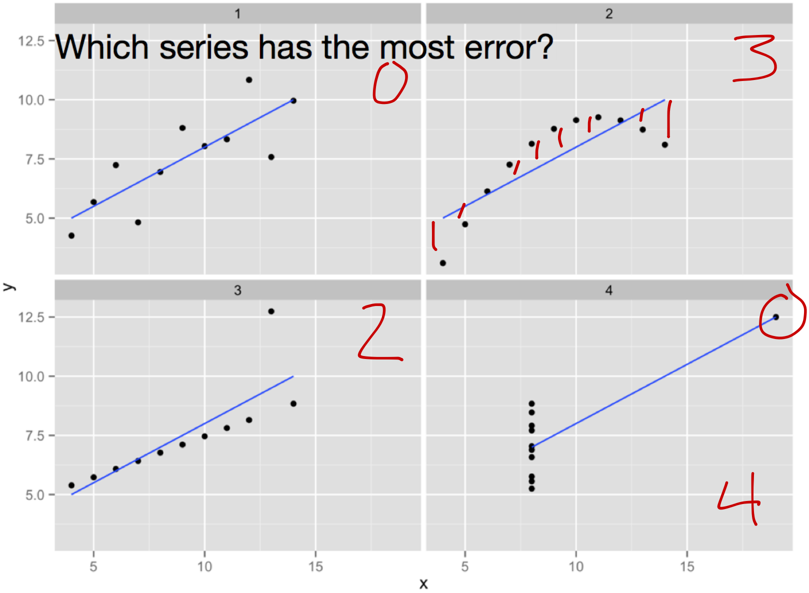
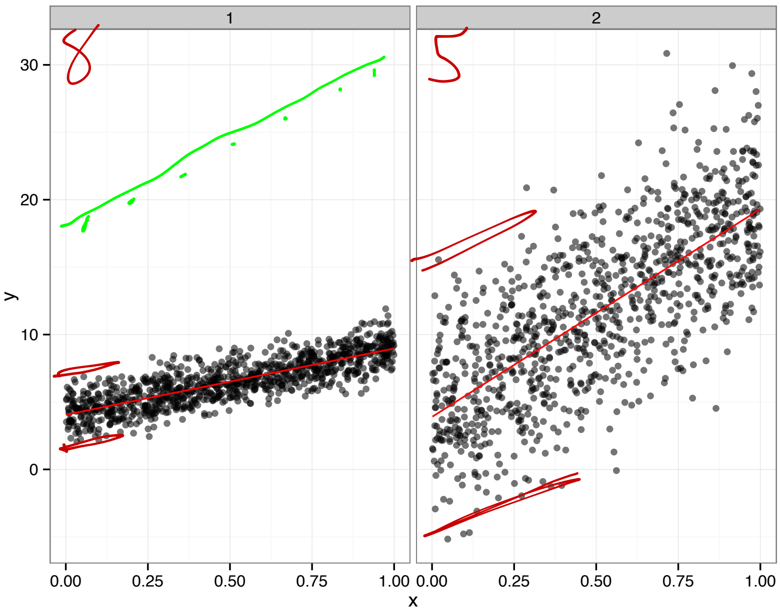


Figure acknowledgements to Hadley Wickham.

Which data show a stronger association?

- less error
-



Goals for this class

You should be able to...

- interpret regression coefficients.
- derive estimators for SLR coefficients.
- implement a SLR from scratch (i.e. not using `lm()`).
- explain why some points have more influence than others on the fitted line.

Regression modeling

- Want to use predictors to learn about the outcome distribution, particularly conditional expected value.
- Formulate the problem parametrically

$$E(y \mid x) = f(x; \beta) = \beta_0 + \beta_1 x_1 + \cancel{\beta_2 x_2 +} \ldots$$

- (Note that other useful quantities, like covariance and correlation, tell you about the joint distribution of y and x)

Brief Detour: Covariance and Correlation

$$
\begin{aligned}
cov(x, y) &= \mathbb{E}\left[(x - \mu_x)(y - \mu_y)\right] \\
cor(x, y) &= \frac{cov(x, y)}{\sqrt{var(x)var(y)}}
\end{aligned}
$$

(units x)

	x	y
1		
2		
⋮		
n		

$$= E\left[xy - x\mu_y - \mu_x y + \mu_x \mu_y\right]$$

$$= E\left[xy\right] - E\left[x\mu_y\right] - \mu_x \mu_y - \mu_x \mu_y + \mu_x \mu_x$$

$$= E(xy) - \mu_x \mu_y$$

$$= \sum_{i=1}^{n} \frac{x_i y_i}{n} - \bar{x}\,\bar{y}$$

Simple linear regression

- Linear models are a special case of all regression models; simple linear regression is the simplest place to start
- Only one predictor:

$x_0 = 1$

$$E(y \mid x) = f(x; \beta) = \beta_0 + \beta_1 x_1$$

X_0 | X_1

- Useful to note that $x_0 = 1$ (implicit definition)
- Somehow, estimate β_0, β_1 using observed data.

Coefficient interpretation

$$E(y|x) = \beta_0 + \beta_1 \cdot x$$

$$E(y|x=0) = \beta_0 + \beta_1 \cdot 0 = \beta_0$$

$$E(y|x=1) = \beta_0 + \beta_1 \cdot 1 = \beta_0 + \beta_1 \cdot k$$

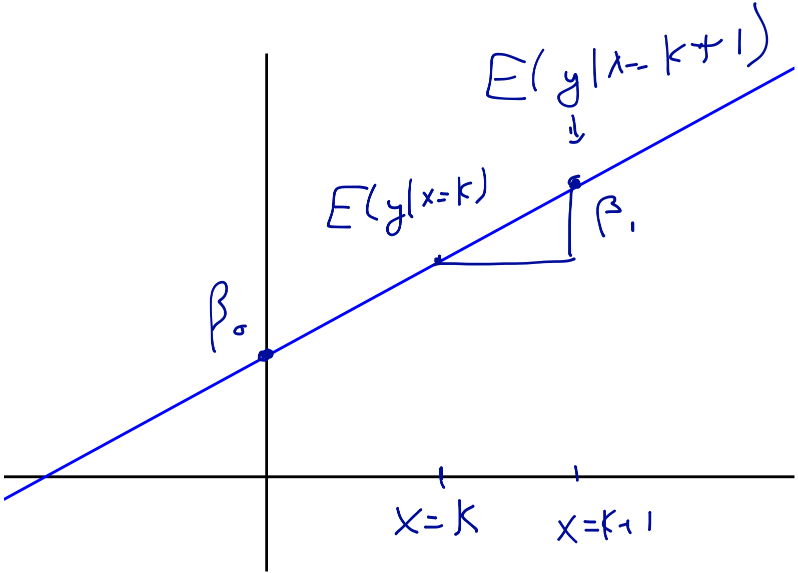
$$- E(y|x=0) = \qquad = \beta_0 + \beta_1 (k-1)$$

$$E(y|x=1) - E(y|x=0) = \beta_1$$

$$E(y|x=k) - E(y|x=k-1) = \beta_1$$

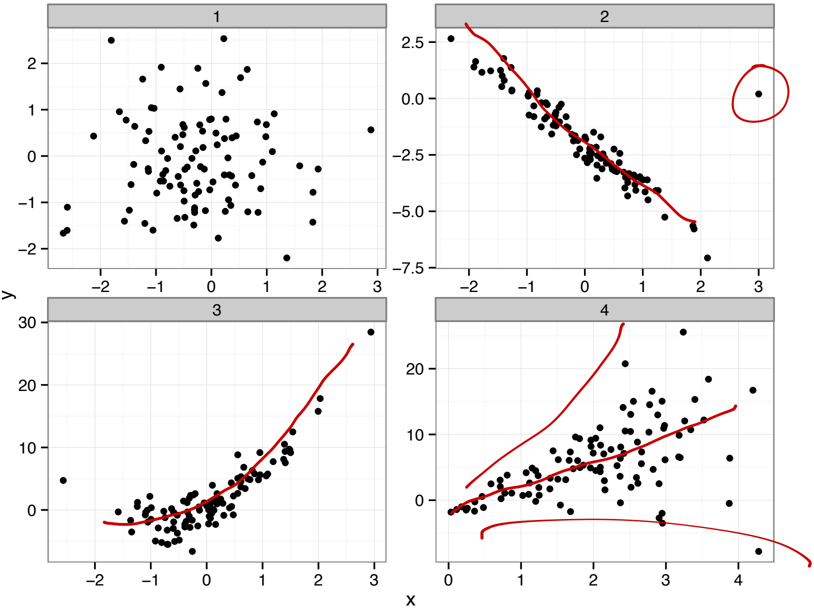
$\beta_1 = \Delta$ in y for a 1-unit $\uparrow$ of x.

Coefficient interpretation



Step 1: Always look at the data!

- Plot the data using, e.g. the `plot()` or `qplot()` functions
- Do the data look like the assumed model?
- Should you be concerned about outliers?
- Define what you expect to see before fitting any model.


1
2
3
4
y
x
2
1
0
−1
−2
2.5
0.0
−2.5
−5.0
−7.5
30
20
10
0
−2
−1
0
1
2
3
4

Least squares estimation

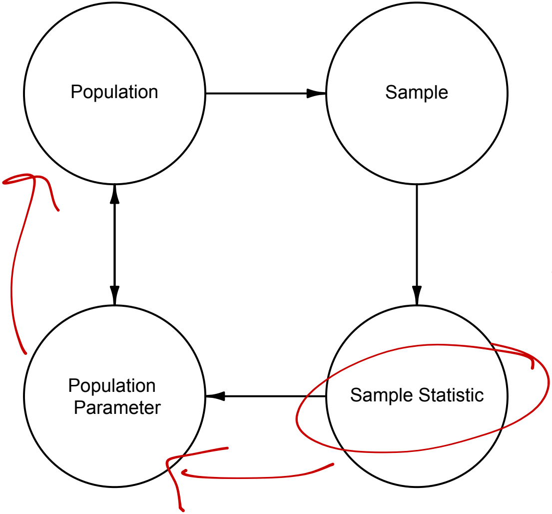
- Observe data (y_i, x_i) for subjects $1, \ldots, I$. Want to estimate β_0, β_1 in the model

$N(0, \sigma^2)$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i; \ \epsilon_i \stackrel{iid}{\sim} (0, \sigma^2)$$

- Recall the assumptions:
  - A2: $E(\epsilon \mid x) = E(\epsilon) = 0$
  - A3: Uncorrelated errors
  - A4: Constant variance
  - A5: Normal distribution not needed for least squares, but **is** needed for inference.]

Circle of Life



$\hat{\beta}_0 \quad \hat{\beta}_1$

Least squares estimation

- Recall that for a single sample $y_i, i \in 1, \ldots, N$, the sample mean $\hat{\mu}_y$ minimizes the sum of squared deviations.

$$RSS(\mu_y) = \sum_{i=1}^{N} (y_i - \mu_y)^2 = (y_1 - \mu_y)^2 + \ldots$$

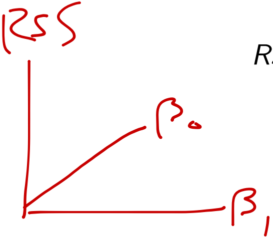
$$\frac{\partial}{\partial \mu} RSS(\mu_y) = -2 \sum_{i=1}^{N} (y_i - \mu_y) \equiv 0$$

$$\sum_{i=1}^{N} y_i - N \cdot \mu_y = 0 \quad /N$$

$$\frac{\sum y_i}{N} = \mu_y$$

Least squares estimation

Find $\hat{\beta}_0$ and β_1. By minimizing RSS relative to each parameter.



$$RSS(\beta_0, \beta_1) = \sum_{i=1}^{N} (y_i - \mathbb{E}[y_i | x_i])^2$$

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

$$\frac{\partial RSS}{\partial \beta_0} =$$

$$\frac{\partial RSS}{\partial \beta_1} =$$

We obtain

$$\hat{\beta}_0 = b_0 = \bar{y} - b_1 \bar{x}$$

$$\hat{\beta}_1 = b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Notes about LSE

Relationship between correlation and slope

$$\rho = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}; \qquad \hat{\beta}_1 = \frac{\widehat{cov}(x,y)}{\widehat{var}(x)} \approx \frac{x \cdot y}{x^2} = \frac{y}{x}$$

Why we need to keep watch for outliers

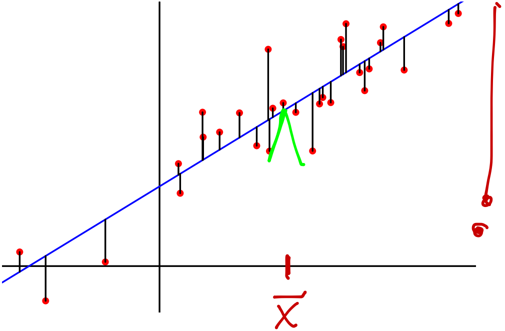
$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \cdot \frac{(x_i - \bar{x})}{(x_i - \bar{x})} \\
&= \frac{\sum \frac{y_i - \bar{y}}{x_i - \bar{x}} (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \rightarrow \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \\
&= \sum \frac{y_i - \bar{y}}{x_i - \bar{x}} \omega_i
\end{aligned}$$

Note that weight ω_i increases as x_i gets further away from $\bar{x}$.

Geometric interpretation of least squares

Least squares minimizes the sum of squared vertical distances between observed and estimated y's:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{I} (y_i - (\beta_0 + \beta_1 x_i))^2$$



Least squares foreshadowing

- Didn't have to choose to minimize squares – could minimize absolute value, for instance.
- Least squares estimates turn out to be a "good idea" – unbiased, BLUE.
- Later we'll see about maximum likelihood as well.