Simple Linear Regression and the Method of Least Squares

Author: Nicholas G Reich, Jeff Goldsmith

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

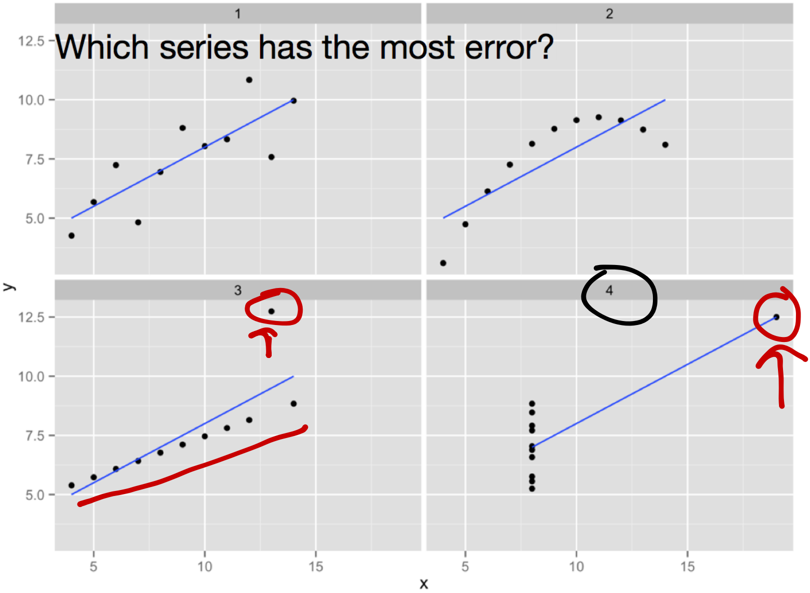
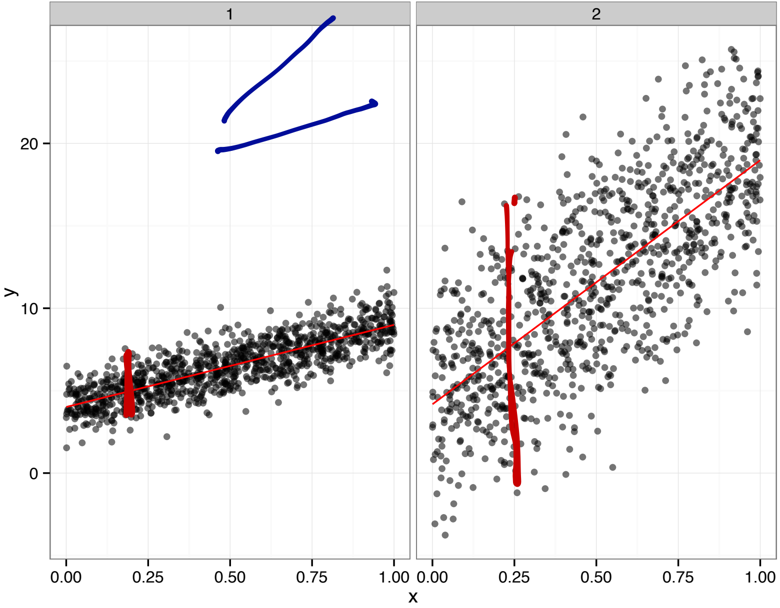


Figure acknowledgements to Hadley Wickham.

Which data show a stronger association?



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

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

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

Brief Detour: Covariance and Correlation

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

$$
\begin{aligned}
cov(x,y) &= E\left[x\cdot y - x\cdot \mu_y - y\cdot \mu_x + \mu_x \mu_y\right] \\
&= E[xy] - \underbrace{E[x]\mu_y}_{\mu_x\cdot\mu_y} - \underline{E[y]\mu_x} + \mu_x \mu_y \\
&= E[xy] - \mu_x \cdot \mu_y \\
&= \sum_{i=1}^{N} \frac{x_i y_i}{N} - \bar{x}\cdot\bar{y}
\end{aligned}
$$

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:

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

SLR

$\Downarrow$

1 x

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

$$E(y|x) = \beta_0 \cdot x_0 + \beta_1 x_1$$

$$x_0 = 1$$

Coefficient interpretation

$$E[y|x] = \beta_0 + \beta_1 \cdot X$$

$$E(y|x=0) = \beta_0$$

$$E[y|x=k] = \cancel{\beta_0} + \cancel{\beta_1 \cdot} k$$

$$E[y|x=k-1] = \beta_0 + \beta_1(k-1)$$

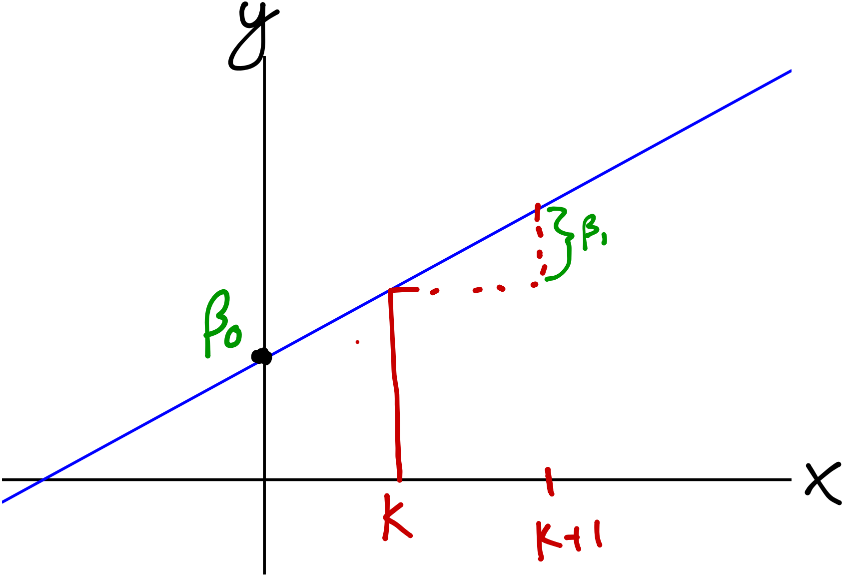
$$= \cancel{\beta_0} + \cancel{\beta_1 \cdot k} - \beta_1$$

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

β_1 = expected change in y for a one unit increase in X

β_0 = expected value of y when $x=0$

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.

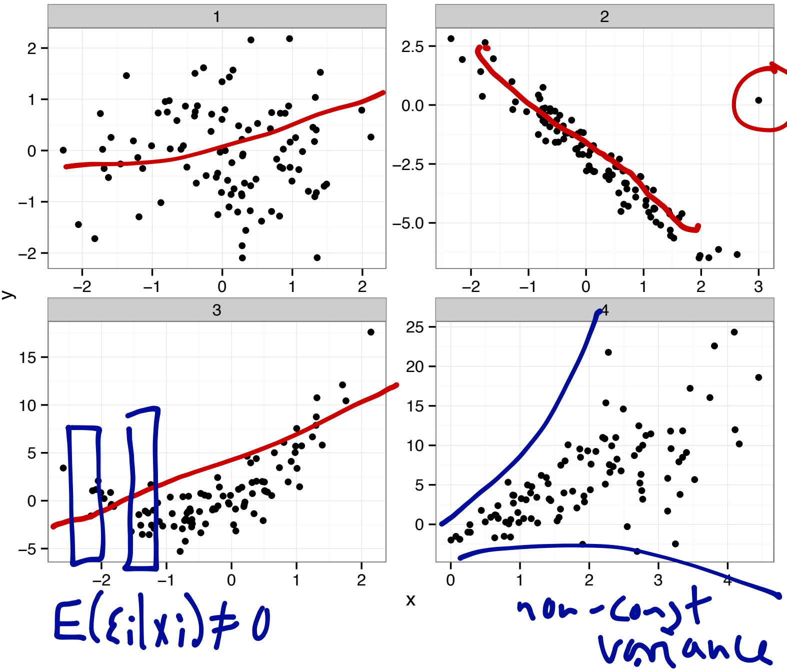
Least squares estimation

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

independent identically distributed

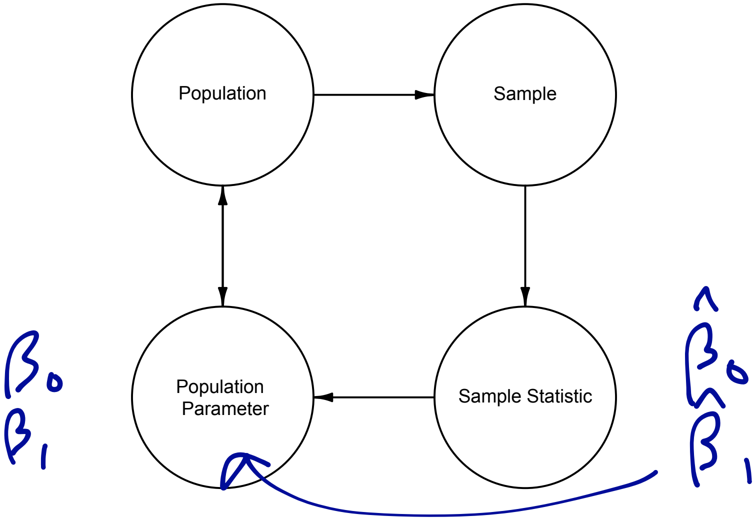
$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i; \; \epsilon_i \overset{iid}{\sim} (0, \sigma^2)$$

- Recall the assumptions:
  - A1: The model: e.g. $y_i = \underline{f(x_i; \beta)} + \epsilon_i = \beta_0 + \beta_1 x_{i,1} + \epsilon_i$
  - A2: Unbiased errors: $\mathbb{E}[\epsilon_i | x_i] = \mathbb{E}[\epsilon_i] = 0$
  - A3: Uncorrelated errors: $cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$.
  - A4: Constant variance: $Var[y_i | x_i] = \sigma^2$
  - A5: Probability distribution: e.g. $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ [not needed for LS, is needed for inference].
  - A6: Representative sampling: generalize to population.



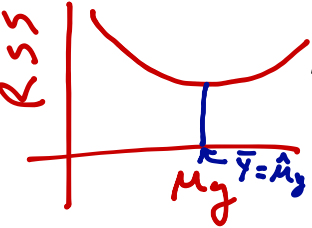
1
2
3
4
y
x
$E(\varepsilon_i|x_i) \neq 0$
non-const variance

Circle of Life



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.



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

$$\frac{\partial RSS(\mu_y)}{\partial \mu_y} = \frac{\partial}{\partial \mu_y}\left[\sum_{1}^{N} (y_i - \mu_y)^2\right]$$

$$= \sum_{i=1}^{N} 2 \cdot (y_i - \mu_y) \cdot -1 \equiv 0$$

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

$$\frac{\sum y_i}{N} = \mu_y \Rightarrow \underline{\hat{\mu}_y = \bar{y}}$$

minimize RSS

Least squares estimation

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

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

$$= \sum \left(y_i - [\beta_0 + \beta_1 x_{1i} \cdots] \right)^2$$

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

$$\rho = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}; \qquad \beta_1 = \frac{cov(x,y)}{var(x)}$$

$$= \frac{cov(x,y)}{sd(x)\ sd(y)} \cdot \frac{sd(y)}{sd(x)} \nearrow$$

Why we need to keep watch for outliers

β_1 vs $\hat{\beta}_1$

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

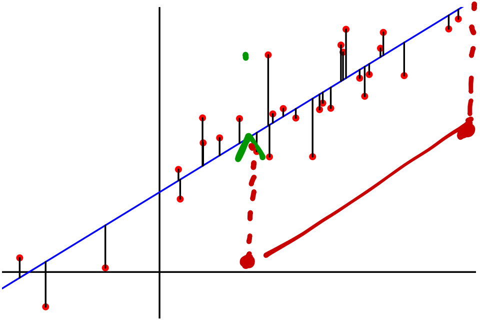
Note that weight ω_i increases as x_i gets further away from $\bar{x}$.

$(y_i - \bar{x}) = 0, \quad \omega_i = 0$

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 (Best Linear Unbiased Estimator).
- Later we'll see about maximum likelihood as well.

Lab exercise: computing $\hat{\beta}$ on your own

- Load the heights data from lecture 1.
- Run a linear model using the R function `lm()`, with daughter height as the outcome.
- Compare the results of that regression with hand-calculated $\hat{\beta}_0$ and $\hat{\beta}_1$ coefficients.

```r
# sample code
install.packages("alr3")
library(alr3)
data(heights)
fm1 <- lm(Dheight ~ Mheight, data=heights)
summary(fm1)
```