Lecture 02
Simple Linear Models: OLS

04 September 2015

Taylor B. Arnold
Yale Statistics
STAT 312/612

Yale

Website

http://euler.stat.yale.edu/~tba3/stat612

Office Hours

- Taylor Arnold:
  - 24 Hillhouse, Office # 206
  - Wednesdays 13:30-15:00, or by appointment
  - Short one-on-one meetings (or small groups)
- Jason Klusowski:
  - 24 Hillhouse, Main Classroom
  - Tuesdays 19:00-20:30
  - Group Q&A style

Simple Linear Models: MLEs

Considering observing n samples from a simple linear model with only a single unknown slope parameter $\beta \in \mathbb{R}$, Considering observing n samples from a simple linear model with only a single unknown slope parameter $\beta \in \mathbb{R}$,

$$y_i = x_i\beta + \epsilon_i, \quad i = 1, \ldots n.$$

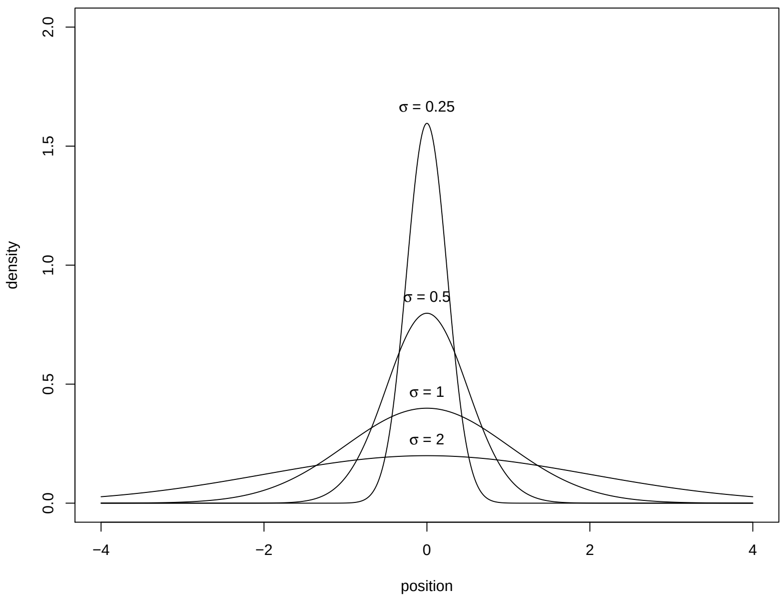
This is, perhaps, the simpliest linear model.

For today, we will assume that the x_i's are fixed and known quantities. This is called a **fixed design**, compared to a **random design**.

The error terms are assumed to be independent and identically distributed random variables with a normal density function:

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

For some unknown variance $\sigma^2 > 0$.



σ = 0.25
σ = 0.5
σ = 1
σ = 2
density
position
0.0
0.5
1.0
1.5
2.0
−4
−2
0
2
4

The density function of a normally distributed random variable with mean μ and variance σ^2 is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \times exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

The density function of a normally distributed random variable with mean μ and variance σ^2 is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \times exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

Conceptually, the front term is just a normalization to make the density sum to 1. The important part is:

$$f(x) \propto exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

Which you have probably seen rewritten as:

$$f(x) \propto exp\left\{-0.5 \cdot \left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

Let's look at the maximum likelihood function of this model:

$$\mathcal{L}(\beta, \sigma | x, y) \quad = \quad \prod_i \mathcal{L}(\beta, \sigma | x_i, y_i)$$

Let's look at the maximum likelihood function of this model:

$$\begin{aligned}\mathcal{L}(\beta, \sigma | x, y) &= \prod_i \mathcal{L}(\beta, \sigma | x_i, y_i) \\ &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \times exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right\}\end{aligned}$$

Let's look at the maximum likelihood function of this model:

$$\begin{aligned}\mathcal{L}(\beta,\sigma|x,y) &= \prod_i \mathcal{L}(\beta,\sigma|x_i,y_i)\\ &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \times exp\left\{-\frac{1}{2\sigma^2}(y_i-\beta x_i)^2\right\}\end{aligned}$$

Notice that the mean μ from the general case has been replaced by βx_i, which should be the mean of $y_i|x_i$.

We can bring the product up into the the exponent as a sum:

$$\mathcal{L}(\beta, \sigma | x, y) \quad = \quad \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \times exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right\}$$

We can bring the product up into the the exponent as a sum:

$$
\begin{aligned}
\mathcal{L}(\beta, \sigma | x, y) &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \times exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right\} \\
&= (2\pi\sigma^2)^{-n/2} \times exp\left\{-\frac{1}{2\sigma^2} \cdot \sum_i (y_i - \beta x_i)^2\right\}
\end{aligned}
$$

Let's highlight the slope parameter β:

$$\mathcal{L}(\beta, \sigma | x, y) \quad = \quad (2\pi\sigma^2)^{-n/2} \times exp\left\{-\frac{1}{2\sigma^2} \cdot \sum_i (y_i - \beta x_i)^2\right\}$$

Let's highlight the slope parameter β:

$$\mathcal{L}(\beta, \sigma | x, y) \quad = \quad (2\pi\sigma^2)^{-n/2} \times exp\left\{ -\frac{1}{2\sigma^2} \cdot \sum_i (y_i - \beta x_i)^2 \right\}$$

What is the MLE for β?

Without resorting to any fancy math, we can see that:

$$\widehat{\beta}_{MLE} = \arg\min_{b \in \mathbb{R}} \left\{ \sum_i (y_i - b \cdot x_i)^2 \right\} \tag{1}$$

The least squares estimator.

A slightly more ‘mathy’ approach would be to calculate the the negative log-likelihood:

A slightly more 'mathy' approach would be to calculate the the negative log-likelihood:

$$-\log\left\{\mathcal{L}(\beta, \sigma | x, y)\right\} = \frac{n}{2} \cdot \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_i (y_i - \beta x_i)^2$$

A slightly more ‘mathy’ approach would be to calculate the the negative log-likelihood:

$$-\log\left\{\mathcal{L}(\beta, \sigma | x, y)\right\} = \frac{n}{2} \cdot \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_i (y_i - \beta x_i)^2$$

Now the minimum of this corrisponds with the maximum likelihood estimators.

Again, we notice that only the second term depends on β:

Again, we notice that only the second term depends on β:

$$-\log\left\{\mathcal{L}(\beta, \sigma | x, y)\right\} = \frac{n}{2} \cdot \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_i (y_i - \beta x_i)^2$$

Again, we notice that only the second term depends on β:

$$-\log\left\{\mathcal{L}(\beta, \sigma | x, y)\right\} = \frac{n}{2} \cdot \log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_i (y_i - \beta x_i)^2$$

And we can again see without resorting to derivatives that the maximum likelihood estimator is that one that minimizes the sum of squares:

$$\widehat{\beta}_{mle} = \underset{b \in \mathbb{R}}{\arg\min} \left\{ \sum_i (y_i - b x_i)^2 \right\}$$

It is possible to directly solve the least squares and obtain an analytic solution to the simple linear regression model.

Taking the derivative of the sum of squares with respect to β we get:

$$\frac{\partial}{\partial\beta}\sum_i (y_i - \beta x_i)^2 \quad = 2 \cdot \textstyle\sum_i (y_i - \beta x_i) \cdot x_i$$

It is possible to directly solve the least squares and obtain an analytic solution to the simple linear regression model.

Taking the derivative of the sum of squares with respect to β we get:

$$\begin{aligned}\frac{\partial}{\partial\beta}\sum_i (y_i - \beta x_i)^2 &= 2 \cdot \textstyle\sum_i (y_i - \beta x_i) \cdot x_i \\ &= 2 \cdot \textstyle\sum_i (y_i x_i - \beta x_i^2)\end{aligned}$$

Setting the derivative equal to zero:

$$2 \cdot \sum_i (y_i x_i - \widehat{\beta} x_i^2) \qquad = 0$$

Setting the derivative equal to zero:

$$
\begin{aligned}
2 \cdot \sum_i (y_i x_i - \widehat{\beta} x_i^2) &= 0 \\
\sum_i y_i x_i &= \widehat{\beta} \textstyle\sum_i x_i^2
\end{aligned}
$$

Setting the derivative equal to zero:

$$
\begin{aligned}
2 \cdot \sum_i (y_i x_i - \widehat{\beta} x_i^2) &= 0 \\
\sum_i y_i x_i &= \widehat{\beta} \sum_i x_i^2 \\
\widehat{\beta}_{MLE} &= \frac{\sum_i y_i x_i}{\sum_i x_i^2}
\end{aligned}
$$

Setting the derivative equal to zero:

$$
\begin{aligned}
2 \cdot \sum_i (y_i x_i - \widehat{\beta} x_i^2) &= 0 \\
\sum_i y_i x_i &= \widehat{\beta} \textstyle\sum_i x_i^2 \\
\widehat{\beta}_{MLE} &= \frac{\sum_i y_i x_i}{\sum_i x_i^2}
\end{aligned}
$$

If you have seen the standard simple least squares solution (that is, with an intercept) this should look familiar.

There are many ways of thinking about the maximum likelihood estimator, one of which is as a weighted sum of the data points y_i:

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum_i y_i x_i}{\sum_i x_i^2} \\
&= \sum_i \left(y_i \cdot \frac{x_i}{\sum_j x_i^2} \right) \\
&= \sum_i y_i w_i
\end{aligned}$$

So, we are weighting the data y_i according to:

$$w_i \propto x_i$$

Does this make sense? **Why**?

One thing that the weighted form of the estimator makes obvious is that the estimator is distributed normally:

$$\widehat{\beta} \sim \mathcal{N}(\cdot, \cdot)$$

As it is the sum of normally distributed variables (y_i).

The mean of the estimator becomes

$$\begin{aligned}\mathbb{E}\widehat{\beta} &= \sum_i \mathbb{E}(y_i w_i) \\ &= \sum_i w_i \cdot \mathbb{E}(y_i) \\ &= \sum_i \beta x_i w_i\end{aligned}$$

The mean of the estimator becomes

$$
\begin{aligned}
\mathbb{E}\widehat{\beta} &= \sum_i \mathbb{E}(y_i w_i) \\
&= \sum_i w_i \cdot \mathbb{E}(y_i) \\
&= \sum_i \beta x_i w_i \\
&= \beta \cdot \sum_i x_i \frac{x_i}{\sum_j x_j^2} \\
&=
\end{aligned}
$$

The mean of the estimator becomes

$$
\begin{aligned}
\mathbb{E}\widehat{\beta} &= \sum_i \mathbb{E}(y_i w_i) \\
&= \sum_i w_i \cdot \mathbb{E}(y_i) \\
&= \sum_i \beta x_i w_i \\
&= \beta \cdot \sum_i x_i \frac{x_i}{\sum_j x_j^2} \\
&= \beta
\end{aligned}
$$

And so we see the estimator is unbiased