Simple Linear Regression

Tanush Chopra

1 Background

To preface this derivation, I first want to discuss the general idea behind regression. The general principle behind regression is that we have this outcome variable, denoted as Y, which we want to model using some explanatory variables x. Note that in this case x is a singular set of explanatory variable values and Y is a singular set of outcome variable values such that x_i corresponds to value Y_i. In this particular derivation, we assume that there exists some linear relationship between x and Y such that for any given datapoint i,

$$\hat{Y}_i = \beta_0 + \beta_1 x_i$$

There are two main reasons why you would want to use a regression model.

1. "Predict" or estimate a future value of Y_i given x_i.
2. "Quantify" the the relationship between Y_i and x_i. This means that for every increase in a unit of x_i you want to find how Y_i changes.

The simplest form of a simple linear regression model just equates the outcome variable Y_i with the linear estimation function, adding error term ϵ_i to model the discrepancy between the $\hat{Y}_i$ and Y_i. This gives us

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

.

2 Normal Distribution

Before discussing Y and x as variables, let's discuss them as random selections along a normal distribution. Specifically, let Y be modeled by $N(\mu, \sigma^2)$. Note there is no dependence on x meaning that we can model Y_i solely using μ and ϵ_i where ϵ_i is the deviation from the mean. This gives us

$$Y_i = \mu + \epsilon_i$$

Given this model of Y_i we can deduce, given n randomly selected points of the distribution $N(\mu, \sigma^2)$, what the mean is given these points. This is given by the function (which represents the likelihood of μ given Y_i and the distribution $N(\mu, \sigma^2)$. This yields

$$\mathcal{L}(\mu|Y) = \prod_{i=1}^{n} P(Y_i|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Y_i-\mu)^2}{2\sigma^2}}$$

Given this, we can now use maximum likelihood estimation to determine a value for μ. This gives us

$$\begin{aligned}
\hat{\mu} = \arg\max_{\mu} \mathcal{L}(\mu|Y) &= \arg\max_{\mu} \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Y_i-\mu)^2}{2\sigma^2}} \\
&= \arg\max_{\mu} \log\big(\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Y_i-\mu)^2}{2\sigma^2}}\big) \\
&= \arg\max_{\mu} \sum_{i=1}^{n} \log(\frac{1}{\sigma\sqrt{2\pi}}) + \log(e^{-\frac{(Y_i-\mu)^2}{2\sigma^2}}) \\
&= \arg\max_{\mu} \sum_{i=1}^{n} \log(e^{-\frac{(Y_i-\mu)^2}{2\sigma^2}}) \\
&= \arg\max_{\mu} \sum_{i=1}^{n} -\frac{(Y_i-\mu)^2}{2\sigma^2} \\
&= \arg\min_{\mu} \sum_{i=1}^{n} (Y_i-\mu)^2
\end{aligned}$$

With this value for $\hat{\mu}$ we can now take the partial derivative with respect to μ and set it to 0 to determine what the expected value of μ is. This gives us

$$\begin{aligned}
\frac{\partial}{\partial\mu} \log \mathcal{L}(\mu|Y) &= 0 \\
\frac{\partial}{\partial\mu} \sum_{i=1}^{n} (Y_i-\mu)^2 &= 0 \\
\sum_{i=1}^{n} -2(Y_i-\mu) &= 0 \\
n\mu &= \sum_{i=1}^{n} Y_i \\
\mu &= \frac{1}{n}\sum_{i=1}^{n} Y_i
\end{aligned}$$

This implies that $E[Y] = \mu$.

Quick Aside

In regression analysis, it's crucial to recognize that Y is a random variable (a function that maps from sample space to measure space), not a deterministic function of x. When correctly modeled, the distribution of Y appears bell-shaped and centers around the mean μ in a histogram. A common misunderstanding arises when the error term ϵ is de-emphasized, leading some to incorrectly assume Y is fixed at μ. Instead, y's values are normally distributed around μ, with μ representing the expected value.

3 Regression

Given that we are now taking Y to be a random variable, we can now model the expected value of Y_i given the explanatory variable x_i. This gives us

$$E[Y_i|x_i] = \beta_0 + \beta_1 x_i$$

Combining this with $E[Y] = \mu$, we have that

$$Y \sim N(\beta_0 + \beta_1 x, \sigma^2)$$

This implies that Y_i is a random normal variable with an expected value of $\beta_0 + \beta_1 x_i$. Given this, we now aim to estimate for β_0 and β_1 parameters.

Using maximum likelihood estimation and the same steps we used to find the estimation of $\hat{\mu}$ we get

$$\begin{aligned}\beta &= \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_i))^2 \\ &= \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2\end{aligned}$$

This gives us an equation that is commonly known as ordinary least squares. Now to solve for β_0 and β_1. Note that $\hat{\beta}_0$ and $\hat{\beta}_1$ represent the estimators for the parameters.

Solving for β_0

The partial derivative of β with respect to β_0 is:

$$\frac{\partial \beta}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)$$

Setting $\frac{\partial \beta}{\partial \beta_0}$ equal to zero gives us

$$-2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0$$

Solving this equation for $\hat{\beta}_0$ gives us

$$\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} Y_i - \frac{\hat{\beta}_1}{n} \sum_{i=1}^{n} x_i = \bar{Y} - \hat{\beta}_1 \bar{x}$$

where $\bar{Y}$ and $\bar{x}$ is the mean of Y and x respectively.

Solving for β_1

Similarly, the partial derivative with respect to β_1 is:

$$\frac{\partial \beta}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (Y_i - \beta_0 - \beta_1 x_i)$$

Setting $\frac{\partial \beta}{\partial \beta_1}$ equal to zero gives us.

$$\begin{aligned}
-2 \sum_{i=1}^{n} x_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\
\sum_{i=1}^{n} x_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 &= 0 \\
\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + \hat{\beta}_0 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} x_i Y_i \\
\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + \left(\bar{Y} - \hat{\beta}_1 \bar{x}\right) \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} x_i Y_i \\
\hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i \right) &= \sum_{i=1}^{n} x_i Y_i - \bar{Y} \sum_{i=1}^{n} x_i \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i Y_i - \bar{Y} \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i}
\end{aligned}$$

The numerator can be expressed as

$$\begin{aligned}
\sum_{i=1}^{n} x_i Y_i - \bar{Y} \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y} \\
&= \sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y} - n\bar{x}\bar{Y} + n\bar{x}\bar{Y} \\
&= \sum_{i=1}^{n} x_i Y_i - \bar{Y} \sum_{i=1}^{n} x_i - \bar{x} \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \bar{x}\bar{Y} \\
&= \sum_{i=1}^{n} \left(x_i Y_i - x_i \bar{Y} - \bar{x} Y_i + \bar{x}\bar{Y} \right) \\
&= \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})
\end{aligned}$$

The denominator can be expressed as

$$
\begin{aligned}
\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}\bar{x} + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}^2 \\
&= \sum_{i=1}^{n} \left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) \\
&= \sum_{i=1}^{n} (x_i - \bar{x})^2
\end{aligned}
$$

Given this we get

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i Y_i - \bar{Y}\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i} \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned}
$$

4 Findings

With all this in mind we find that

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1\bar{x} \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned}
$$

5 Additional Problems

1. Find $E[\hat{\beta}_0]$
2. Find $V[\hat{\beta}_0]$
3. Find $E[\hat{\beta}_1]$
4. Find $V[\hat{\beta}_1]$
5. What do you notice about $\hat{\beta}_0$ and $\hat{\beta}_1$ based on there expected values? Are the estimators biased? (Hint: I'm BLUE da ba dee da ba di)