Linear Regression Introduction

Ciencia de Datos para la toma de decisiones en Economía

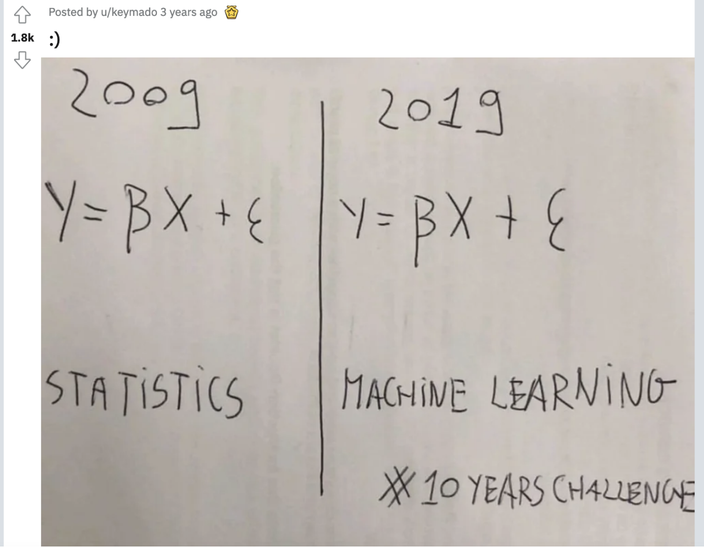
Ignacio Sarmiento-Barbieri

Agenda

Intro

- Linear regression is the “work horse” of econometrics and (supervised) machine learning.
- Very powerful in many contexts.
- Big ‘payday’ to study this model in detail.

Intro


Posted by u/keymado 3 years ago
1.8k
:)
2009
Y = βX + ε
STATISTICS
2019
Y = βX + ε
MACHINE LEARNING
10 YEARS CHALLENGE

Linear Regression Model

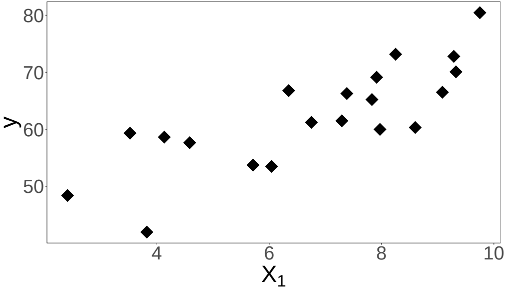
- If $f(X) = X\beta$, obtaining $f(.)$ boils down to obtaining β

$$y = X\beta + u \tag{1}$$

- where
  - y is a vector $n \times 1$ with typical element y_i
  - X is a matrix $n \times k$
    - Note that we can represent it as a column vector $\underset{n\times k}{X} = [\underset{n\times 1}{X_1}\ \underset{n\times 1}{X_2} \dots \underset{n\times 1}{X_k}]$
  - β is a vector $k \times 1$ with typical element β_j

Linear Regression Model

For example, we have the following data and want to estimate a linear model:



We would specify:

$$y = \beta_0 + \beta_1 X_1 + u \tag{2}$$

Linear Regression Model

- How do we obtain β?
  - Method of Moments H.W.
  - MLE (more on this later)
  - OLS: minimize SSR ($e'e$)
    - where $e = Y - \hat{Y} = Y - X\hat{\beta}$

OLS

How do we obtain β?

- Consider the following risk function, where we minimize the sum of square residuals

$$SSR(\tilde{\beta}) \equiv \sum_{i=1}^{n} \tilde{e}_i^2 = \tilde{e}'\tilde{e} = (y - X\tilde{\beta})'(y - X\tilde{\beta}) \tag{3}$$

  - $SSR(\tilde{\beta})$ is the aggregation of squared errors if we choose $\tilde{\beta}$ as an estimator.

- The **least squares estimator** $\hat{\beta}$ will be

$$\hat{\beta} = \underset{\tilde{\beta}}{argmin}\, SSR(\tilde{\beta}) \tag{4}$$

OLS

$$SSR(\tilde{\beta}) = \tilde{e}'\tilde{e} \tag{5}$$
$$= (Y - X\tilde{\beta})'(y - X\tilde{\beta}) \tag{6}$$

- FOC are

$$\frac{\partial \tilde{e}'\tilde{e}}{\partial \tilde{\beta}} = 0 \tag{7}$$
$$\frac{\partial \tilde{e}'\tilde{e}}{\partial \tilde{\beta}} = -2X'y + 2X'X\tilde{\beta} = 0 \tag{8}$$

OLS

- Let $\hat{\beta}$ be the solution.Then $\hat{\beta}$ satisfies the following normal equation

$$X'X\hat{\beta} = X'y \tag{9}$$

- If the inverse of $X'X$ exists, then

$$\hat{\beta} = (X'X)^{-1}X'y \tag{10}$$

- Note that this is a closed solution (a bonus!!)

Assessing the Accuracy of the Coefficient Estimates

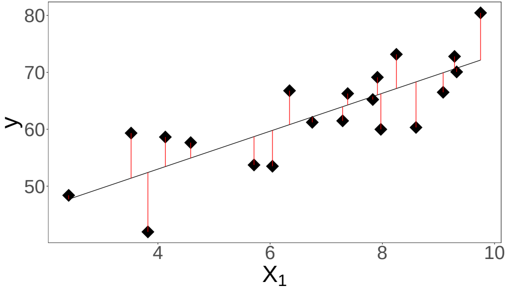
- **Predicting well in this context $\rightarrow$ estimating well**. Why?

Assessing the Accuracy of the Coefficient Estimates

- **Predicting well in this context** $\rightarrow$ **estimating well**. Why?
  - The prediction of y will be given by $\hat{y} = X\hat{\beta}$

Assessing the Accuracy of the Coefficient Estimates

- **Predicting well in this context → estimating well**. Why?
  - The prediction of y will be given by $\hat{y} = X\hat{\beta}$



  - In our simple example the prediction of y will be given by $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$
  - A natural question is how accurate are $\hat{\beta}_0$ and $\hat{\beta}_1$ as an estimate of β_0 and β_1?

Assessing the Accuracy of the Coefficient Estimates

- The variance of $\hat{\beta}_0$

$$Var(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right] \tag{11}$$

- and $\hat{\beta}_1$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{12}$$

Agenda

Statistical Properties

Under certain assumptions HW Review the Assumption from Econometrics

- Small Sample (Gauss-Markov Theorem)
  - Unbiased: $E(\hat{\beta}) = \beta$
  - Minimum Variance: $Var(\tilde{\beta}) - Var(\hat{\beta})$ is positive semidefinite matrix Proof: HW. Remember: a matrix $M_{p\times p}$ is positive semi-definite iff $c'Mc \geq 0 \ \forall c \in \mathbb{R}^p$
- Large Sample
  - Consistency: $\hat{\beta} \to_p \beta$
  - Asymptotically Normal: $\sqrt{N}(\hat{\beta} - \beta) \sim_a N(0, S)$

Gauss Markov Theorem

- Gauss Markov Theorem that says OLS is BLUE is perhaps one of the most famous results in statistics.
  - $E(\hat{\beta}) = \beta$
  - $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$
- and implies that $\hat{y}$ is an unbiased predictor and minimum variance, from the class of unbiased linear predictors (BLUP) H.W. proof

Gauss Markov Theorem

- Gauss Markov Theorem that says OLS is BLUE is perhaps one of the most famous results in statistics.
  - $E(\hat{\beta}) = \beta$
  - $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$
- and implies that $\hat{y}$ is an unbiased predictor and minimum variance, from the class of unbiased linear predictors (BLUP) H.W. proof
- However, it is essential to note the limitations of the theorem.
  - Correctly specified with exogenous Xs,
  - The term error is homoscedastic
  - No serial correlation.
  - Nothing about the OLS estimator being the more efficient than any other estimator one can imagine.

Asymptotic Tests

- Continuing with our simple example

$$y = \beta_0 + \beta_1 X_1 + u \tag{13}$$

- Suppose that you want to test if there's is no relationship between X_1 and y
- The large sample properties + the assumptions needed to get there allows us to have valid tests

Asymptotic Tests

- Mathematically testing that there's is no relationship between X_1 and y corresponds to testing

$$H_0 = \beta_1 = 0 \tag{14}$$

$$H_1 = \beta_1 \neq 0 \tag{15}$$

- and use a $t - statistic$

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} \tag{16}$$

Asymptotic Tests

- However, in some cases the asymptotic approximation need not be very good
- Especially with highly nonlinear models
- Resampling methods can allow us to somewhat quantify uncertainty and improve on the asymptotic distribution approximations
- Bootstrapping, which is a popular resampling method, can be used as an alternative to asymptotic approximations for obtaining standard errors, confidence intervals, and p-values for test statistics.

Further Readings

- Davidson, R., & MacKinnon, J. G. (2004). Econometric theory and methods (Vol. 5). New York: Oxford University Press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.