Multiple Linear Equation Models (Take 2)

Ethan Ligon

March 12, 2024

The Fundamental Linear Regression Model

Start with

$$y = \boldsymbol{X}\beta + u.$$

- Allowing $(y, \boldsymbol{X}, u)$ to all be random.
- $\boldsymbol{X}$ has full column rank.
- Key identification assumption: $\mathbb{E}(X^\top u) = \mathbf{0}$.

Compare Classical Approach

E.g., R.A. Fisher; (Fisher Box, 1980)

$$y = \boldsymbol{X}\beta + u.$$

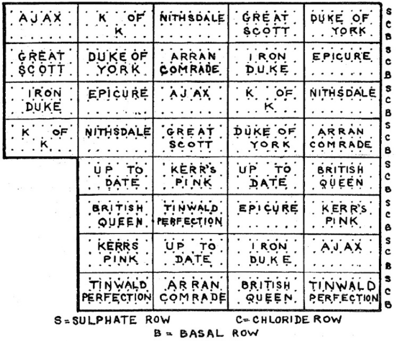
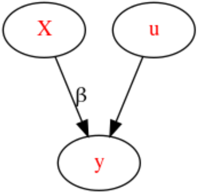


Diagram 1. Plan of experiment. Farmyard manure series.

Figure: "Triplicate Chessboard": Diagram I from Fisher and Mackenzie (1923)

Classical Interpretation

The "dependent" variable y is determined by some random variables X with observations realized, and some random unobserved u. Critically, u and X are orthogonal; i.e., $\mathbb{E}(X^\top u) = \mathbf{0}$.



- The orthogonality of X and u is *not testable*, since u isn't observed.
- In general the causal diagram above imposes needed structure for interpreting regression.
- With this structure, β is "effect of variation in X on y."

Classical Regression in python

See `classical_regression.ipynb` on `datahub`.

Compare Bayes

$$y = \boldsymbol{X}\beta + u.$$

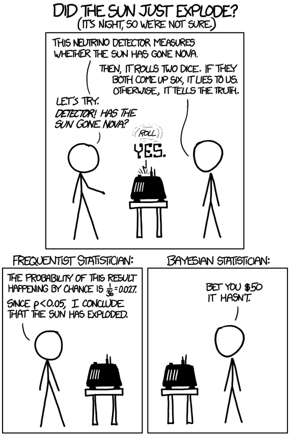
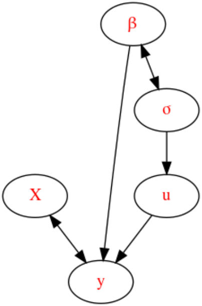


Figure: https://xkcd.com/1132/

The disturbance u is independently distributed $Q(u|\sigma)$, and there's a "prior" $\Pr(\beta = \beta, \sigma = \sigma) = \Pr(\beta, \sigma)$ over these unknown vectors. In addition the variables (β, σ) are ordinarily assumed to be distributed independently of $\boldsymbol{X}$.



Some Linear Estimation Problems with Multiple Equations

Let us develop some results for a broad class of linear estimators. We've seen special cases of this before, so should look familiar. Suppose we have an random vector $\boldsymbol{T}$.

Linear Model

$$y = \boldsymbol{X}\beta + u.$$

Premultiply by $\boldsymbol{T}$

$$\boldsymbol{T}^\top y = \boldsymbol{T}^\top \boldsymbol{X}\beta + \boldsymbol{T}^\top u.$$

We want to solve this equation for β. How should we proceed?

DISCUSS

Pretend you don't know any econometrics or statistics. What are issues? What if $\boldsymbol{T}$ is just a column of ones?

Aside on Moore-Penrose Inverse

https://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_inverse
For any real matrix $\boldsymbol{A}$ (need not be square!), let $\boldsymbol{A}^+$ satisfy:

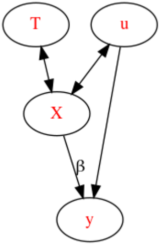
1. $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}$; (generalizes idea that $\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{I}$)
2. $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}^+$;
3. $(\boldsymbol{A}^+\boldsymbol{A})^\top = \boldsymbol{A}\boldsymbol{A}^+$; (form of symmetry)
4. $(\boldsymbol{A}\boldsymbol{A}^+)^\top = \boldsymbol{A}^+\boldsymbol{A}$.

Any such $\boldsymbol{A}^+$ satisfying these conditions is called the "Moore-Penrose Inverse" (or sometimes the pseudo-inverse).
Facts about the Moore-Penrose Inverse:

1. $\boldsymbol{A}^+$ exists and is unique.
2. If the columns of $\boldsymbol{A}$ are linearly independent, then $\boldsymbol{A}^+ = (\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top$ (this is sometimes called the "left inverse), and $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}$.
3. If the rows of $\boldsymbol{A}$ are linearly independent, then $\boldsymbol{A}^+ = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}$, and $\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{I}$.

A$_1$ $\mathbb{E}\boldsymbol{T}^\top u = \mathbf{0}$ (Orthogonality)

A$_2$ Some kind of full rank condition on $\boldsymbol{T}^\top \boldsymbol{X}$ (e.g., full column rank with probability one).



- The orthogonality of $\boldsymbol{T}$ and u is *not testable*, since u isn't observed.
- In general the causal diagram above imposes needed structure for interpreting regression.
- With this structure, β is "effect of variation in X on y."

Least Squares Estimator

Now:

$$\boldsymbol{T}^\top y = \boldsymbol{T}^\top \boldsymbol{X}\beta + \boldsymbol{T}^\top u$$
$$\Rightarrow \boldsymbol{T}^\top y - \boldsymbol{T}^\top u = (\boldsymbol{T}^\top \boldsymbol{X})\beta.$$

Now, using A_2:

$$(\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top y) - (\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top u) = \beta.$$

We only get a sample of realizations $(\boldsymbol{y}, \boldsymbol{X})$, and never even observe realizations of u. So to make further progress we take expectations and exploit A_1:

$$\mathbb{E}(\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top y) = \beta.$$

Analogy Principle (or Monte Carlo integration)

We have

$$\mathbb{E}(\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top y) = \beta.$$

We then apply the analogy principal, which allows us to substitute the mean of a sample for the expected value, yielding

$$b^* = (\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top \boldsymbol{Y}),$$

Then let we have $\mathbb{E}b^* = \beta$; i.e., this procedure results in an unbiased estimator of β for any $\boldsymbol{T}$ satisfying A_1 & A_2.

Linear Weighted Regression in `python`

See `weighted_regression.ipynb` on `datahub`.