Multiple Regressions

Data-Based Economics

Year 2022-2023

The problem

Remember dataset from last time

	type	income	education	prestige
accountant	prof	62	86	82
pilot	prof	72	76	83
architect	prof	75	92	90
author	prof	55	90	76
chemist	prof	64	86	90

- Last week we “ran” a linear regression: $y = \alpha + \beta x$. Result:

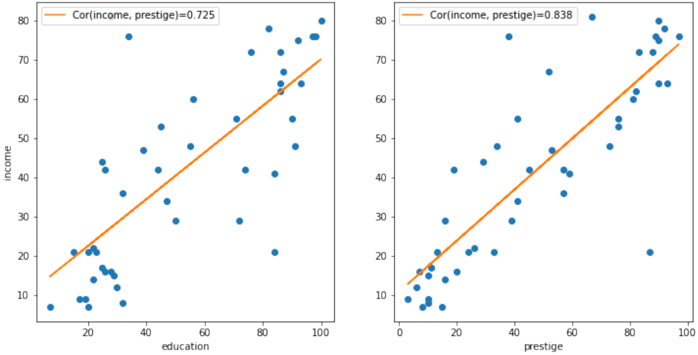
$$\text{income} = xx + 0.72\text{education}$$

- Should we have looked at “prestige” instead ?

$$\text{income} = xx + 0.83\text{prestige}$$

- Which one is better?

Prestige or Education



- if the goal is to predict: the one with higher explained variance
  - `prestige` has higher R^2 (0.83^2)
- unless we *are* interested in the effect of education

Multiple regression

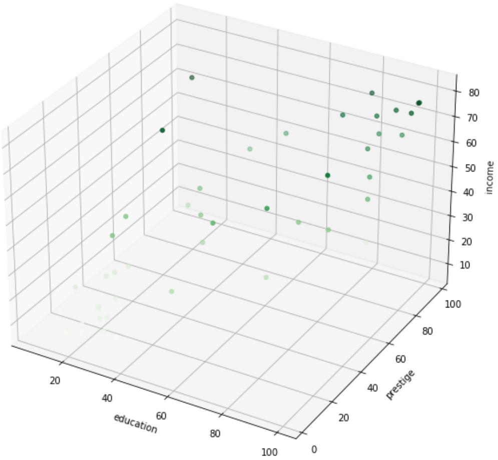
- What about using both?
  - 2 variables model:

$$\text{income} = \alpha + \beta_1 \text{education} + \beta_2 \text{prestige}$$

  - will *probably* improve prediction power (explained variance)
  - β_1 might not be meaningful on its own anymore (education and prestige are correlated)

Fitting a model

Now we are trying to fit a **plane** to a cloud of points.



Minimization Criterium

- Take all observations: $(\text{income\_}n, \text{education\_}n, \text{prestige\_}n)\_n \in [0, N]$
- Objective: sum of squares

$$L(\alpha, \beta_1, \beta_2) = \sum_i \Big(\underbrace{\alpha + \beta_1 \text{education\_}n + \beta_2 \text{prestige\_}n - \text{income\_}n}\_e_n = \text{prediction}$$

- Minimize loss function in α, β_1, β_2
- Again, we can perform numerical optimization (machine learning approach)
  - ... but there is an explicit formula

Ordinary Least Square

$$Y = \begin{bmatrix} \text{income}_1 \\ \vdots \\ \text{income}_N \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & \text{education}_1 & \text{prestige}_1 \\ \vdots & \vdots & \vdots \\ 1 & \text{education}_N & \text{prestige}_N \end{bmatrix}$$

- Matrix Version (look for $B = (\alpha, \beta_1, \beta_2)$):

$$Y = XB + E$$

- Note that constant can be interpreted as a "variable"
- Loss function

$$L(A, B) = (Y - XB)'(Y - XB)$$

- Result of minimization $\min_{(A,B)} L(A, B)$:

$$\begin{bmatrix} \alpha & \beta_1 & \beta_2 \end{bmatrix} = (X'X)^{-1}X'Y$$

Solution

- Result:

$$\text{income} = 10.43 + 0.03 \times \text{education} + 0.62 \times \text{prestige}$$

Explained Variance