Multiple Linear Regression

Tyler Ransom

Univ of Oklahoma

Jan 29, 2019

Today's plan

1. Review reading topics
   1.1 Multiple Regression Model
   1.2 Formulas for OLS Estimates
   1.3 Interpretation
2. In-class activity: Practice running multiple linear regressions

The Multiple Regression Model

Multiple Regression: What & Why

- **multiple regression model** $=$ simple regression model, but with 3+ variables
- Reasons for having more than two independent variables:
  1. Better chance of uncovering causal effects
  2. Can build better models for predicting y
  3. Can incorporate more flexible functional form relationships

Motivation for Multiple Regression

- Consider an extension of the $\log(wage)$ equation we used for simple regression:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + u$$

  where *IQ* is IQ score (in the population, it has a mean of 100 and sd = 15)

- Primarily interested in β_1, but β_2 is of some interest, too.

- *IQ* included in the equation $\Rightarrow$ no longer in error term

- If *IQ* proxies intelligence, gets us closer to causal effect of schooling

The model with two x's

- Generally, we can write a model with two independent variables as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

  where:

- β_0 is the intercept
- β_1 measures the change in y with respect to x_1, **holding other factors fixed**
- β_2 measures the change in y with respect to x_2, **holding other factors fixed**

The key assumption

- As before, the key assumption about how u is related to the x's is

$$E(u|x_1, x_2) = 0$$

- For any values of x_1 and x_2 in the population, the average unobservable $= 0$
- In the wage equation, the assumption is $E(u|educ, IQ) = 0$
- Now u no longer contains intelligence (we hope), and so this condition has a better chance of being true.
- In simple regression, we had to assume IQ and $educ$ are unrelated to justify leaving IQ in the error term.

The model with $K > 1$ x's

- The **multiple linear regression model** can be written in the population as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

  where β_0 is the **intercept**, β_1 is the parameter associated with x_1, β_2 is the parameter associated with x_2, and so on.

- Contains $k + 1$ (unknown) population parameters. We call β_1, ..., β_k the **slope parameters**.

- Now we have multiple explanatory or independent variables. We still have one explained or dependent variable. We still have an error term, u.

More flexible functional forms

- Consider the following regression model:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exper + \beta_4 exper^2 + u,$$

  where *exper* is labor market experience (in years)

- Take $x_1 = educ$, $x_2 = IQ$, $x_3 = exper$ and $x_4 = exper^2$

- Note that x_4 is a **nonlinear** function of x_3

- From last time: $100\beta_1 \approx \%\Delta wage$ when $educ \uparrow$ one year

- $100\beta_2$ has a similar interpretation (for a one point increase in IQ).

- What about β_3 and β_4?

Quadratic forms

- β_3 and β_4 are harder to interpret

- can use calculus to get the slope of $\log(wage)$ with respect to *exper*:

$$\frac{\partial \log(wage)}{\partial exper} = \beta_3 + 2\beta_4 exper$$

- multiply by 100 to get the percentage effect

- More on this later in the course

The key assumption with $K > 1$ x's

- The assumption that $E(u|x_1, x_2) = 0$ simply becomes

$$E(u|x_1, x_2, \ldots, x_K) = 0$$

  when we have more than two x's

- Can make this condition more plausible by "controlling for" more variables

- In the wage example, control for IQ when estimating return to education

Deriving multivariate OLS formulas

Key assumptions

- Just like before, make use of these two assumptions:

$$
\begin{aligned}
E(u) &= 0 \\
Cov(x_1, u) = E(x_1 u) &= 0 \\
Cov(x_2, u) = E(x_2 u) &= 0 \\
\vdots &= \vdots \\
Cov(x_k, u) = E(x_k u) &= 0
\end{aligned}
$$

- Second through last lines implied by "$E(u|x_1, \ldots, x_k) = E(u)$ for all x"
    - In other words, if u and x_k are mean independent, then their covariance $= 0$
- If it helps: x is now a $N \times K$ **matrix** instead of a $N \times 1$ vector

Plugging the model into the assumption formulas

- Now let's plug in our multiple regression model for u in the prior formulas:

$$
\begin{aligned}
E(y - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k) &= 0 \\
E[x_1(y - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k)] &= 0 \\
\vdots &= \vdots \\
E[x_k(y - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k)] &= 0
\end{aligned}
$$

- Gives us $k+1$ equations and $k+1$ unknowns
- This set of formulas is known as the **OLS first order conditions**
- Solution will be the OLS estimates of $\beta_0, \ldots, \beta_k$

Interpretation

Interpretation of the $\hat{\beta}$'s

- The slope coefficients now explicitly have **ceteris paribus** interpretations

- Consider $k = 2$:
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$
Then
$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2$$
allows us to compute how predicted y changes when x_1 and x_2 change by any amount

- If we "hold x_2 fixed," i.e. its change is zero, $\Delta x_2 = 0$
$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 \text{ if } \Delta x_2 = 0$$

Interpretation of the $\hat{\beta}$'s

- Similarly,

$$\Delta\hat{y} = \hat{\beta}_2 \Delta x_2 \text{ if } \Delta x_1 = 0$$

  and

$$\hat{\beta}_2 = \frac{\Delta\hat{y}}{\Delta x_2} \text{ if } \Delta x_1 = 0$$

- We call $\hat{\beta}_1$ and $\hat{\beta}_2$ **partial effects** or **ceteris paribus effects**

Meaning of *ceteris paribus*

- What does it mean to “hold other factors fixed”?
- multiple regression:
    - gives us the ceteris paribus interpretation
    - without us having to find two people with the same value of *IQ* who differ in education by one year
- The estimation method does that for us. It is able to do it because we assume a particular relationship, in this case

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + u$$

- keep in mind that this is due to the **linear in parameters** assumption
- and the **unconfoundedness** assumption ($E(u|x_1, \ldots, x_k) = 0$)

Fitted values and residuals

- We have the same definitions as before for fitted values & residuals
- Same for SSE, SSR, SST, and R^2
- And the SRF goes through the point $(\bar{x}_1, \ldots, \bar{x}_k, \bar{y})$
- Thing to know:
  - using the same set of data and the same dependent variable,
  - the R-squared can **never fall** when another x is added to the model

Frisch-Waugh theorem

- One formula for $\hat{\beta}_1$ is
$$\hat{\beta}_1 = \frac{\sum_{i=1}^N \hat{r}_{i1} y_i}{\sum_{i=1}^N \hat{r}_{i1}^2}$$
where $\hat{r}_{i1}$ is the residual from a regression of x_1 on $x_2, \ldots, x_k$

- Recall that the residuals are the part of y that's uncorrelated with the x's

- In this case, x_1 is the dependent variable

- So β_1 is the part of x_1 that's uncorrelated with $x_2, \ldots, x_k$ but corr. with y

- Gives another interpretation of partial effect

Comparing Simple and Multiple Regression Estimates

- Consider the simple and multiple OLS regression lines:

$$\begin{aligned}\tilde{y} &= \tilde{\beta}_0 + \tilde{\beta}_1 x_1 \\ \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2\end{aligned}$$

  where a tilde ($\sim$) denotes the simple regression and hat ($\hat{}$) denotes multiple regression (using the same N observations).

- Question: Is there a simple relationship between $\tilde{\beta}_1$ (which does not control for x_2) and $\hat{\beta}_1$ (which does)?

- Yes, but we need to define another simple regression. (See next slide)

Comparing Simple and Multiple Regression Estimates

- Let $\widetilde{\delta}_1$ be the slope from the regression

$$x_2 = \delta_1 x_1 + u_x$$

- It is always true for any sample that

$$\widetilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \widetilde{\delta}_1$$

- Notice: slope coefficient on x_2 in the multiple regression ($\hat{\beta}_2$) also plays a role. (See next slide)

Comparing Simple and Multiple Regression Estimates

- **Case 1:** If the partial effect of x_2 on y is positive, so $\hat{\beta}_2 > 0$, and x_1 and x_2 are positive correlated in the sample, so $\widetilde{\delta}_1 > 0$, then

$$\widetilde{\beta}_1 > \hat{\beta}_1$$

- **Case 2:** If $\hat{\beta}_2 > 0$ and $\widetilde{\delta}_1 < 0$ (x_1 and x_2 negatively correlated), then

$$\begin{aligned}\widetilde{\beta}_1 =& \hat{\beta}_1 + \hat{\beta}_2\widetilde{\delta}_1 \\ =& \hat{\beta}_1 + (+)(-) \\ =& \hat{\beta}_1 + (-) \\ <& \hat{\beta}_1\end{aligned}$$

The multiple regression plane

- With 2+ x-variables in the regression, the OLS line becomes a **plane**

- **`Interactive graphic`** (about two-thirds of the way down the page)