Regression III

Lecture 7

Quantitative Methods for Political Economy
7SSPN107

Recap: simple regression and OLS

Last lecture's example: how is education X_i related to hourly earnings Y_i?

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = -6 + 2X_i.$$

One more education year is associated with £2/hour higher earnings.

Recap: simple regression and OLS

Last lecture's example: how is education X_i related to hourly earnings Y_i?

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = -6 + 2X_i.$$

One more education year is associated with £2/hour higher earnings.

At $X_i = 14$, predicted earnings are $-6 + 2(14) = 22$ £/hour.

If actual earnings are £25/hour, the residual is $\hat{u}_i = 25 - 22 = +3$.

Recap: simple regression and OLS

Last lecture's example: how is education X_i related to hourly earnings Y_i?

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = -6 + 2X_i.$$

One more education year is associated with £2/hour higher earnings.

At $X_i = 14$, predicted earnings are $-6 + 2(14) = 22$ £/hour.

If actual earnings are £25/hour, the residual is $\hat{u}_i = 25 - 22 = +3$.

OLS chooses the coefficients to minimise the sum of squared residuals, $\sum_{i=1}^{n} \hat{u}_i^2$.

Recap: hypothesis tests and confidence intervals

The education coefficient is 2. Suppose that the robust standard error is 0.8.

$$H_0 : \beta_1 = 0, \qquad t = \frac{2.00 - 0}{0.80} = 2.50.$$

Recap: hypothesis tests and confidence intervals

The education coefficient is 2. Suppose that the robust standard error is 0.8.

$$H_0 : \beta_1 = 0, \qquad t = \frac{2.00 - 0}{0.80} = 2.50.$$

For a two-sided 5% test in a large sample, we reject H_0 if $|t| > 1.96$, ($p \approx 0.012$).

Recap: hypothesis tests and confidence intervals

The education coefficient is 2. Suppose that the robust standard error is 0.8.

$$H_0 : \beta_1 = 0, \qquad t = \frac{2.00 - 0}{0.80} = 2.50.$$

For a two-sided 5% test in a large sample, we reject H_0 if $|t| > 1.96$, ($p \approx 0.012$).

The 95% confidence interval is

$$2.00 \pm 1.96(0.80) = [0.43,\ 3.57].$$

Note that the interval excludes zero, consistent with the decision to reject H_0.

Recap: a binary regressor

A binary regressors takes only two values.

Let $D_i = 1$ for training participants and $D_i = 0$ for non-participants.

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 D_i = 18 + 3.20 D_i \qquad \text{(earnings in £/hour)}.$$

What are the two group means?

Recap: a binary regressor

A binary regressors takes only two values.

Let $D_i = 1$ for training participants and $D_i = 0$ for non-participants.

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 D_i = 18 + 3.20 D_i \qquad \text{(earnings in £/hour)}.$$

What are the two group means?

- Non-participants, $D_i = 0$: £18.00, the intercept $\hat{\beta}_0$.
- Participants, $D_i = 1$: $18 + 3.20 =$ £21.20, $\hat{\beta}_0 + \hat{\beta}_1$.
- Difference: £3.20, the slope $\hat{\beta}_1$.

Recap: a binary regressor

A binary regressors takes only two values.

Let $D_i = 1$ for training participants and $D_i = 0$ for non-participants.

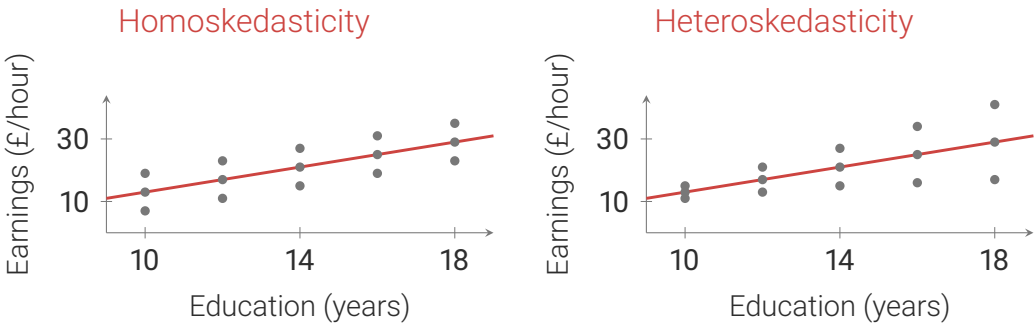
$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 D_i = 18 + 3.20 D_i \qquad \text{(earnings in £/hour)}.$$

What are the two group means?

- Non-participants, $D_i = 0$: £18.00, the intercept $\hat{\beta}_0$.
- Participants, $D_i = 1$: $18 + 3.20 =$ £21.20, $\hat{\beta}_0 + \hat{\beta}_1$.
- Difference: £3.20, the slope $\hat{\beta}_1$.

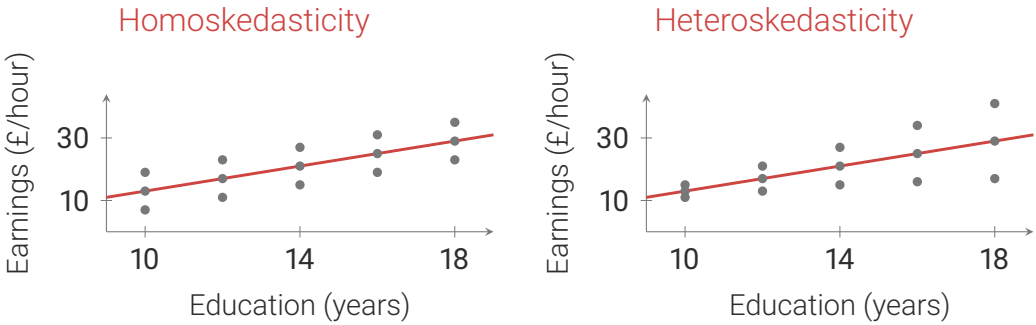
Non-participants are the reference group, $D_i = 0$. Changing the reference group can change the sign of the coefficient, but it does not change its (non-)significance.

Recap: heteroskedasticity and robust standard errors



Heteroskedasticity refers to the case where the spread of observations around the trend line is not constant.

Recap: heteroskedasticity and robust standard errors



Heteroskedasticity refers to the case where the spread of observations around the trend line is not constant.

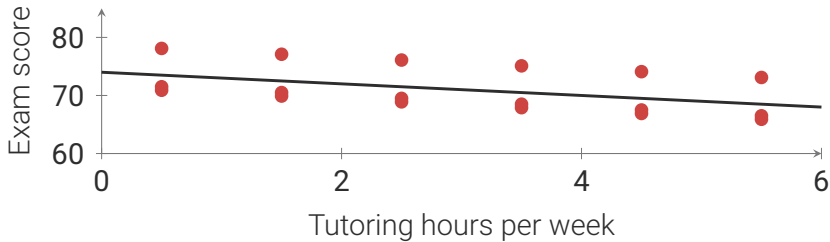
Even with heteroskedasticity, OLS remains unbiased and consistent under the least-squares assumptions, but we need to use robust standard errors for inference in large samples.

6.1–6.2

Why multiple regression?

Tutoring and exam scores

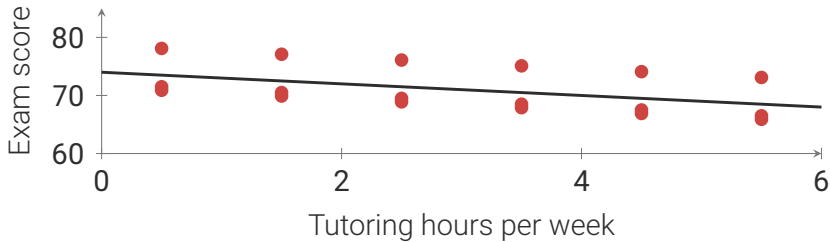
Suppose that we have data on students' exam scores and the number of hours of tutoring that they have received.



Does this mean tutoring makes students perform worse in exams?

Tutoring and exam scores

Suppose that we have data on students' exam scores and the number of hours of tutoring that they have received.



Does this mean tutoring makes students perform worse in exams?

Not necessarily: students with lower previous scores may receive more tutoring.

The multiple regression model

The multiple regression model acknowledges the fact that a variable may be related to more than one regressor.

$$E(Y_i \mid X_{1i}, \ldots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

In the case of exam scores and tutoring, we could also include previous scores:

$$\widehat{\text{Score}}_i = 20 + 5H_i + 0.6P_i.$$

H_i: tutoring hours per week. P_i: previous exam score.

The multiple regression model

The multiple regression model acknowledges the fact that a variable may be related to more than one regressor.

$$E(Y_i \mid X_{1i}, \ldots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

In the case of exam scores and tutoring, we could also include previous scores:

$$\widehat{\text{Score}}_i = 20 + 5H_i + 0.6P_i.$$

H_i: tutoring hours per week. P_i: previous exam score.

One more hour of tutoring is associated with a 5-point higher predicted score, holding previous scores constant.

The multiple regression model

The multiple regression model acknowledges the fact that a variable may be related to more than one regressor.

$$E(Y_i \mid X_{1i}, \ldots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

In the case of exam scores and tutoring, we could also include previous scores:

$$\widehat{\text{Score}}_i = 20 + 5H_i + 0.6P_i.$$

H_i: tutoring hours per week. P_i: previous exam score.

One more hour of tutoring is associated with a 5-point higher predicted score, holding previous scores constant.

One more previous-score point is associated with a 0.6-point higher predicted score, holding tutoring constant.

Exercise 1: comparing students

$$\widehat{\text{Score}}_i = 20 + 5H_i + 0.6P_i.$$

1. Predict the score for a student with 4 tutoring hours and a previous score of 60.

2. The coefficient on tutoring hours is larger than that on previous test scores. Does that mean that tutoring is more important than previous test scores in explaining current test scores?

Exercise 1: comparing students

$$\widehat{\text{Score}}_i = 20 + 5H_i + 0.6P_i.$$

1. Predict the score for a student with 4 tutoring hours and a previous score of 60.

$20 + 5(4) + 0.6(60) = 76$ points.

2. The coefficient on tutoring hours is larger than that on previous test scores. Does that mean that tutoring is more important than previous test scores in explaining current test scores?

Exercise 1: comparing students

$$\widehat{\text{Score}}_i = 20 + 5H_i + 0.6P_i.$$

1. Predict the score for a student with 4 tutoring hours and a previous score of 60.

$20 + 5(4) + 0.6(60) = 76$ points.

2. The coefficient on tutoring hours is larger than that on previous test scores. Does that mean that tutoring is more important than previous test scores in explaining current test scores?

No. The regressors use different units. We also do not know how much variation there is in tutoring hours or previous test scores.

6.3

The OLS estimator in multiple regression

Ordinary least squares

OLS chooses all coefficients together to minimise the sum of squared residuals across the sample.

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki} \right)^2.$$

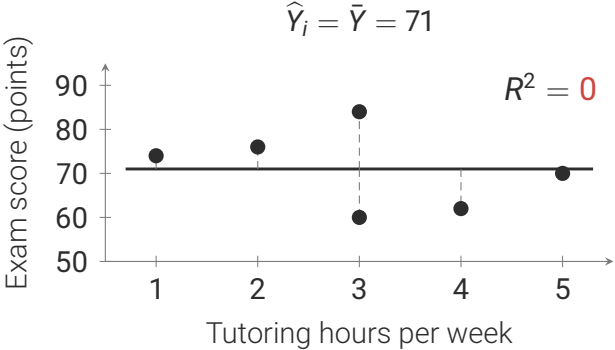
Recall $\hat{u}_i = Y_i - \widehat{Y}_i$.

6.4

How well does the model fit?

R-squared

R^2 measures the proportion of variation in Y around its sample mean that is explained by the fitted model.



$$R^2 = 1 - \frac{\text{SSR}}{\text{TSS}}$$

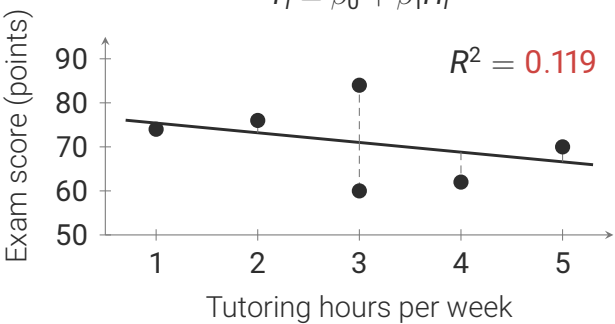
$$\text{SSR} = \sum_i \hat{u}_i^2 \quad = \sum_i (Y_i - \widehat{Y}_i)^2$$

$$\text{TSS} = \sum_i (Y_i - \bar{Y})^2$$

R-squared

R^2 measures the proportion of variation in Y around its sample mean that is explained by the fitted model.

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 H_i$$



$$R^2 = 1 - \frac{\text{SSR}}{\text{TSS}}$$

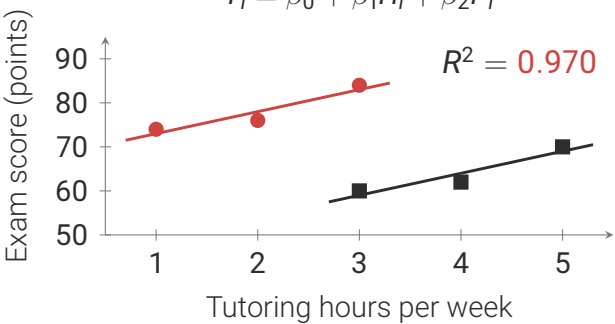
$$\text{SSR} = \sum_i \hat{u}_i^2 \quad = \sum_i (Y_i - \widehat{Y}_i)^2$$

$$\text{TSS} = \sum_i (Y_i - \bar{Y})^2$$

R-squared

R^2 measures the proportion of variation in Y around its sample mean that is explained by the fitted model.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 H_i + \hat{\beta}_2 P_i$$



$$R^2 = 1 - \frac{\text{SSR}}{\text{TSS}}$$

$$\text{SSR} = \sum_i \hat{u}_i^2 = \sum_i (Y_i - \hat{Y}_i)^2$$

$$\text{TSS} = \sum_i (Y_i - \bar{Y})^2$$

Previous score: 80 or 40.

Both fitted lines have slope $+5$.

Adjusted R-squared

OLS minimises the sum of squared residuals. If we add a regressor, the sum of squared residuals would remain the same if the OLS coefficient on the new regressor is zero.

For the same outcome and observations adding regressors cannot lower R^2.

Adjusted R^2 includes a penalty for including additional regressors:

$$\bar{R}^2 = 1 - \frac{\mathrm{SSR}/(n-k-1)}{\mathrm{TSS}/(n-1)},$$

where n is the sample size and k counts regressors, excluding the intercept.

Adjusted R-squared

OLS minimises the sum of squared residuals. If we add a regressor, the sum of squared residuals would remain the same if the OLS coefficient on the new regressor is zero.

For the same outcome and observations adding regressors cannot lower R^2.

Adjusted R^2 includes a penalty for including additional regressors:

$$\bar{R}^2 = 1 - \frac{\mathrm{SSR}/(n-k-1)}{\mathrm{TSS}/(n-1)},$$

where n is the sample size and k counts regressors, excluding the intercept.

Adjusted R^2 can fall if an extra predictor adds too little improvement.

Exercise 2: calculating and interpreting model fit

A hypothetical exam-score regression has an intercept and two regressors:

$$n = 21, \qquad k = 2, \qquad \mathrm{SSR} = 180, \qquad \mathrm{TSS} = 2000.$$

1. Calculate R^2 and adjusted R^2. Remember that k excludes the intercept.

2. Does this high R^2 establish that tutoring causes higher exam scores?

Exercise 2: calculating and interpreting model fit

A hypothetical exam-score regression has an intercept and two regressors:

$$n = 21, \qquad k = 2, \qquad \mathrm{SSR} = 180, \qquad \mathrm{TSS} = 2000.$$

1. Calculate R^2 and adjusted R^2. Remember that k excludes the intercept.

$$R^2 = 1 - \frac{180}{2000} = 0.91,$$
$$\bar{R}^2 = 1 - \frac{180/(21-2-1)}{2000/(21-1)} = 1 - \frac{10}{100} = 0.90.$$

2. Does this high R^2 establish that tutoring causes higher exam scores?

Exercise 2: calculating and interpreting model fit

A hypothetical exam-score regression has an intercept and two regressors:

$$n = 21, \qquad k = 2, \qquad \mathrm{SSR} = 180, \qquad \mathrm{TSS} = 2000.$$

1. Calculate R^2 and adjusted R^2. Remember that k excludes the intercept.

$$R^2 = 1 - \frac{180}{2000} = 0.91,$$
$$\bar{R}^2 = 1 - \frac{180/(21-2-1)}{2000/(21-1)} = 1 - \frac{10}{100} = 0.90.$$

2. Does this high R^2 establish that tutoring causes higher exam scores?

No. R^2 describes fit in the sample. Other differences between students may influence both tutoring and scores.

6.5–6.7

Assumptions and multicollinearity

The least-squares assumptions in multiple regression

1. Zero conditional mean: unexplained influences average to zero at each combination of predictors: $E(u_i \mid X_{1i}, \ldots, X_{ki}) = 0$.

The least-squares assumptions in multiple regression

1. Zero conditional mean: unexplained influences average to zero at each combination of predictors: $E(u_i \mid X_{1i}, \ldots, X_{ki}) = 0$.
2. Independent random observations: observations are independent draws from the same population.

The least-squares assumptions in multiple regression

1. Zero conditional mean: unexplained influences average to zero at each combination of predictors: $E(u_i \mid X_{1i}, \ldots, X_{ki}) = 0$.
2. Independent random observations: observations are independent draws from the same population.
3. Very large outliers are unlikely: extreme values of the regressors and the outcome are sufficiently rare.

The least-squares assumptions in multiple regression

1. Zero conditional mean: unexplained influences average to zero at each combination of predictors: $E(u_i \mid X_{1i}, \ldots, X_{ki}) = 0$.
2. Independent random observations: observations are independent draws from the same population.
3. Very large outliers are unlikely: extreme values of the regressors and the outcome are sufficiently rare.
4. No perfect multicollinearity: predictors must not contain exactly the same information.

The least-squares assumptions in multiple regression

1. Zero conditional mean: unexplained influences average to zero at each combination of predictors: $E(u_i \mid X_{1i}, \ldots, X_{ki}) = 0$.
2. Independent random observations: observations are independent draws from the same population.
3. Very large outliers are unlikely: extreme values of the regressors and the outcome are sufficiently rare.
4. No perfect multicollinearity: predictors must not contain exactly the same information.

Under these assumptions, the OLS estimators are unbiased and consistent. In large samples, they are approximately jointly normally distributed. Each estimator has

$$\hat{\beta}_j \overset{\text{approx.}}{\sim} N\left(\beta_j, \sigma^2_{\hat{\beta}_j}\right), \qquad j = 0, \ldots, k.$$

Perfect and imperfect multicollinearity

Perfect multicollinearity refers to the case where two regressors contain exactly the same information: tutoring hours and tutoring minutes per week. OLS cannot compute separate coefficients for hours and minutes.

Perfect and imperfect multicollinearity

Perfect multicollinearity refers to the case where two regressors contain exactly the same information: tutoring hours and tutoring minutes per week. OLS cannot compute separate coefficients for hours and minutes.

Imperfect multicollinearity refers to the case where two or more regressors are highly but not perfectly correlated.

Perfect and imperfect multicollinearity

Perfect multicollinearity refers to the case where two regressors contain exactly the same information: tutoring hours and tutoring minutes per week. OLS cannot compute separate coefficients for hours and minutes.

Imperfect multicollinearity refers to the case where two or more regressors are highly but not perfectly correlated.

The coefficients can be estimated, but their estimated associations with the dependent variable may be imprecise.

6.1 and 6.8

Omitted variables and controls

Omitted-variable bias

Suppose that the full model of test scores includes tutoring H_i and previous scores P_i:

$$Y_i = \beta_0 + \beta_1 H_i + \beta_2 P_i + u_i, \qquad E(u_i \mid H_i, P_i) = 0.$$

Omitted-variable bias

Suppose that the full model of test scores includes tutoring H_i and previous scores P_i:

$$Y_i = \beta_0 + \beta_1 H_i + \beta_2 P_i + u_i, \qquad E(u_i \mid H_i, P_i) = 0.$$

If we omit P_i, the tutoring slope is (population version):

$$\begin{aligned}
\beta_1^{\text{short}} &= \frac{\text{Cov}(H, Y)}{\text{Var}(H)} = \frac{\text{Cov}(H, \beta_0 + \beta_1 H + \beta_2 P + u)}{\text{Var}(H)} \\
&= \frac{\text{Cov}(H, \beta_0) + \beta_1 \text{Cov}(H, H) + \beta_2 \text{Cov}(H, P) + \text{Cov}(H, u)}{\text{Var}(H)} \\
&= \beta_1 + \underbrace{\beta_2 \frac{\text{Cov}(H, P)}{\text{Var}(H)}}_{\text{omitted-variable bias}}.
\end{aligned}$$

When does an omitted variable cause bias?

In our model, the bias in the tutoring slope is

$$\beta_2 \frac{\text{Cov}(H, P)}{\text{Var}(H)}.$$

There is an omitted variable bias if:

1. $\beta_2 \neq 0$: previous attainment helps determine exam scores, even after accounting for tutoring.

When does an omitted variable cause bias?

In our model, the bias in the tutoring slope is

$$\beta_2 \frac{\text{Cov}(H, P)}{\text{Var}(H)}.$$

There is an omitted variable bias if:

1. $\beta_2 \neq 0$: previous attainment helps determine exam scores, even after accounting for tutoring.
2. $\text{Cov}(H, P) \neq 0$: previous attainment is correlated with tutoring hours.

Example: omitted-variable bias

Suppose the hypothetical population values are:

- Tutoring coefficient in the full model: $\beta_1 = 5$.
- Previous-score coefficient in the full model: $\beta_2 = 0.6$.
- Each additional tutoring hour is associated with a previous score that is 10 points lower: $\mathrm{Cov}(H, P) / \mathrm{Var}(H) = -10$.

Example: omitted-variable bias

Suppose the hypothetical population values are:

- Tutoring coefficient in the full model: $\beta_1 = 5$.
- Previous-score coefficient in the full model: $\beta_2 = 0.6$.
- Each additional tutoring hour is associated with a previous score that is 10 points lower: $\mathrm{Cov}(H, P) / \mathrm{Var}(H) = -10$.

Omitting previous attainment gives a negative bias:

$$\text{Bias} = 0.6 \times (-10) = -6.$$

Example: omitted-variable bias

Suppose the hypothetical population values are:

- Tutoring coefficient in the full model: $\beta_1 = 5$.
- Previous-score coefficient in the full model: $\beta_2 = 0.6$.
- Each additional tutoring hour is associated with a previous score that is 10 points lower: $\mathrm{Cov}(H, P) / \mathrm{Var}(H) = -10$.

Omitting previous attainment gives a negative bias:

$$\text{Bias} = 0.6 \times (-10) = -6.$$

The tutoring slope when previous attainment is omitted is

$$\beta_1^{\text{short}} = 5 - 6 = -1.$$

The negative tutoring slope reflects the lower previous attainment of students receiving more tutoring.

Exercise 3: the direction of bias

Suppose higher previous attainment raises exam scores, even at the same tutoring level.

1. Suppose higher-attaining students receive more tutoring. Which way does omitting previous attainment bias the tutoring slope?

2. Suppose tutoring hours are randomly assigned, so they are unrelated to previous attainment in the population. Would omitting previous attainment bias the tutoring coefficient?

Exercise 3: the direction of bias

Suppose higher previous attainment raises exam scores, even at the same tutoring level.

1. Suppose higher-attaining students receive more tutoring. Which way does omitting previous attainment bias the tutoring slope?

Upward: tutoring also picks up their attainment advantage. Both β_2 and $\text{Cov}(H, P)$ are positive.

2. Suppose tutoring hours are randomly assigned, so they are unrelated to previous attainment in the population. Would omitting previous attainment bias the tutoring coefficient?

Exercise 3: the direction of bias

Suppose higher previous attainment raises exam scores, even at the same tutoring level.

1. Suppose higher-attaining students receive more tutoring. Which way does omitting previous attainment bias the tutoring slope?

Upward: tutoring also picks up their attainment advantage. Both β_2 and $\mathrm{Cov}(H, P)$ are positive.

2. Suppose tutoring hours are randomly assigned, so they are unrelated to previous attainment in the population. Would omitting previous attainment bias the tutoring coefficient?

No bias from omitting previous attainment, because $\mathrm{Cov}(H, P) = 0$.

Exercise 4: apartment rents

Supopse we are interested in the effect of location on apartment rent. We regress monthly apartment rent on distance from the city centre, omitting apartment size.

1. What is the direction of omitted-variable bias in the distance coefficient?

2. If the distance coefficient is negative when we control for floor area, must it become positive when floor area is omitted?

Exercise 4: apartment rents

Supopse we are interested in the effect of location on apartment rent. We regress monthly apartment rent on distance from the city centre, omitting apartment size.

1. What is the direction of omitted-variable bias in the distance coefficient?

Upward: Apartment size likely raises rent and is positively correlated with distance. Both terms in the bias formula are positive.

2. If the distance coefficient is negative when we control for floor area, must it become positive when floor area is omitted?

Exercise 4: apartment rents

Supopse we are interested in the effect of location on apartment rent. We regress monthly apartment rent on distance from the city centre, omitting apartment size.

1. What is the direction of omitted-variable bias in the distance coefficient?

Upward: Apartment size likely raises rent and is positively correlated with distance. Both terms in the bias formula are positive.

2. If the distance coefficient is negative when we control for floor area, must it become positive when floor area is omitted?

No. The omitted variable bias shifts the coefficient upwards, but we don't know if the shift is large enough to turn the coefficient positive.

7.1

Inference in multiple regression

A coefficient and its standard error

As before, the predicted test score is given by:

$$\widehat{\text{Score}} = 20 + 5H + 0.6P.$$

Suppose the tutoring coefficient has a robust standard error of 2.

How do we test the hypothesis that the coefficient on tutoring hours is zero?

Example: inference with a control

Just like before, we construct the t-statistic and its confidence interval.

$$t = \frac{5 - 0}{2} = 2.50, \qquad 95\%\ \text{CI} = 5 \pm 1.96(2) = [1.08,\ 8.92].$$

Example: inference with a control

Just like before, we construct the t-statistic and its confidence interval.

$$t = \frac{5 - 0}{2} = 2.50, \qquad 95\%\ \text{CI} = 5 \pm 1.96(2) = [1.08,\ 8.92].$$

Using the large-sample normal approximation, the two-sided p-value is $p = 2\Pr(Z > 2.50) \approx 0.0124$, where $Z \sim N(0, 1)$.

Example: inference with a control

Just like before, we construct the t-statistic and its confidence interval.

$$t = \frac{5 - 0}{2} = 2.50, \qquad 95\%\ \text{CI} = 5 \pm 1.96(2) = [1.08,\ 8.92].$$

Using the large-sample normal approximation, the two-sided p-value is $p = 2\Pr(Z > 2.50) \approx 0.0124$, where $Z \sim N(0, 1)$.

If we use a 5% significance level, we reject the zero-coefficient null.

Exercise 5: inference about apartment rents

We regress monthly rent on distance from the city centre (km) and floor area (m^2). The distance coefficient is -30, with a robust standard error of 20.

1. Test $H_0 : \beta_{\text{distance}} = 0$ against a two-sided alternative at 5%.

2. Calculate a 95% confidence interval. Does the result establish that distance has no relationship with rent?

Exercise 5: inference about apartment rents

We regress monthly rent on distance from the city centre (km) and floor area (m^2). The distance coefficient is -30, with a robust standard error of 20.

1. Test $H_0 : \beta_{\text{distance}} = 0$ against a two-sided alternative at 5%.

$$t = \frac{-30 - 0}{20} = -1.50.$$

Since $|t| < 1.96$, do not reject the null.

2. Calculate a 95% confidence interval. Does the result establish that distance has no relationship with rent?

Exercise 5: inference about apartment rents

We regress monthly rent on distance from the city centre (km) and floor area (m^2). The distance coefficient is -30, with a robust standard error of 20.

1. Test $H_0 : \beta_{\text{distance}} = 0$ against a two-sided alternative at 5%.

$$t = \frac{-30 - 0}{20} = -1.50.$$

Since $|t| < 1.96$, do not reject the null.

2. Calculate a 95% confidence interval. Does the result establish that distance has no relationship with rent?

$$-30 \pm 1.96(20) = [-69.20,\ 9.20].$$

The interval includes zero, but does not prove there is no relationship.

8.1–8.2

Quadratic regression models

Curved relationships within multiple regression

A straight line assumes that one more unit of ***X*** has the same association with ***Y*** at every value of ***X***.

Curved relationships within multiple regression

A straight line assumes that one more unit of X has the same association with Y at every value of X.

We can relax this assumption by including transformed regressors:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i.$$

Curved relationships within multiple regression

A straight line assumes that one more unit of X has the same association with Y at every value of X.

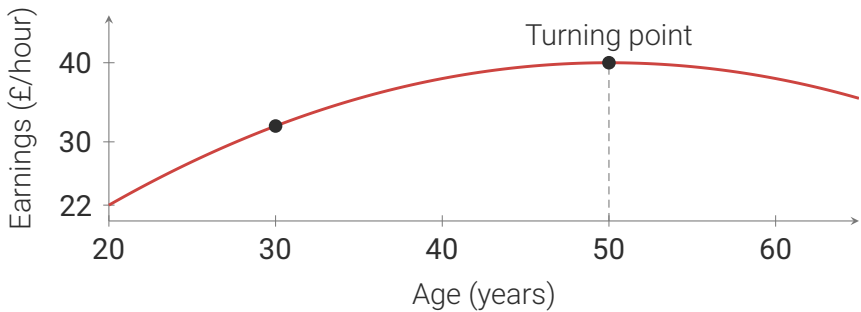
We can relax this assumption by including transformed regressors:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i.$$

The regression is curved in X, but linear in the coefficients. We can still estimate it by OLS.

A quadratic earnings profile

Suppose hourly earnings and age A follow $\widehat{Y} = -10 + 2A - 0.02A^2$.



Interpreting a quadratic

For $\widehat{Y} = -10 + 2A - 0.02A^2$, the slope at age A is

$$\frac{d\widehat{Y}}{dA} = 2 - 0.04A.$$

Interpreting a quadratic

For $\widehat{Y} = -10 + 2A - 0.02A^2$, the slope at age A is

$$\frac{d\widehat{Y}}{dA} = 2 - 0.04A.$$

At age 30 the slope is 0.80. At age 50 it is zero.

Interpreting a quadratic

For $\widehat{Y} = -10 + 2A - 0.02A^2$, the slope at age A is

$$\frac{d\widehat{Y}}{dA} = 2 - 0.04A.$$

At age 30 the slope is 0.80. At age 50 it is zero.

For a full one-year change, compare predictions:

$$\widehat{Y}(A = 31) - \widehat{Y}(A = 30) = 32.78 - 32.00 = 0.78.$$

Exercise 6: age and earnings

Use $\widehat{Y} = -10 + 2A - 0.02A^2$.

1. Calculate predicted earnings at ages 40 and 45. What is the difference?

2. Which null hypothesis tests whether the quadratic term is needed?

Exercise 6: age and earnings

Use $\widehat{Y} = -10 + 2A - 0.02A^2$.

1. Calculate predicted earnings at ages 40 and 45. What is the difference?

$$\widehat{Y}(A = 40) = 38.00, \qquad \widehat{Y}(A = 45) = 39.50.$$

The five-year difference is £1.50/hour.

2. Which null hypothesis tests whether the quadratic term is needed?

Exercise 6: age and earnings

Use $\widehat{Y} = -10 + 2A - 0.02A^2$.

1. Calculate predicted earnings at ages 40 and 45. What is the difference?

$$\widehat{Y}(A = 40) = 38.00, \qquad \widehat{Y}(A = 45) = 39.50.$$

The five-year difference is £1.50/hour.

2. Which null hypothesis tests whether the quadratic term is needed?

$H_0 : \beta_2 = 0$. Use a robust t-test on the coefficient of the squared term.