Applied Microeconometrics (L2): Basic Regression Tools

Nicholas Giannakopoulos

University of Patras
Department of Economics

ngias@upatras.gr

October 22, 2019

Overview

Empirical problem: Class size and educational output

- What?
  - What is the effect of reducing class size by one student per class?
- Why?
  - Economic rationale? Smaller classes promote student learning (Educational production function, Angrist and Lavy, 1999)
- How?
  - Secondary school micro-data on students' achievements and class size
  - Model: $y_{isc} = X_s'\beta + n_{sc}\alpha + \pi_c + \eta_s + \epsilon_{isc}$
  - i: student ($i = 1, ..., N$), c: class ($c = 1, ..., C$), s: school ($s = 1, ..., S$)
  - y: pupil's test score
  - X: school characteristics
  - n: size of class
  - π: random class attributes (i.i.d.)
  - ϵ: disturbance term
  - Other dep vars: parent satisfaction, student personal development, future adult welfare and/or earnings, performance on standardized tests

Case study: The California Test Score Data Set

- All K-6 and K-8 California school districts (n = 420)
- Relationship of interest (Dependent and Independent Variables)
  - 5^{th} grade test scores (Stanford-9 achievement test, combined math and reading), district average
  - Student-teacher ratio (STR) = no. of students in the district divided by no. full-time equivalent teachers

TABLE 4.1 Summary of the Distribution of Student-Teacher Ratios and Fifth-Grade Test Scores for 420 K–8 Districts in California in 1998

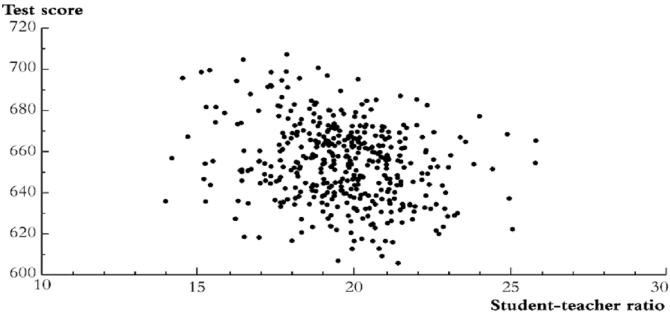
			Percentile						
	Average	Standard Deviation	10%	25%	40%	50% (median)	60%	75%	90%
Student-teacher ratio	19.6	1.9	17.3	18.6	19.3	19.7	20.1	20.9	21.9
Test score	654.2	19.1	630.4	640.0	649.1	654.5	659.4	666.7	679.1

Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

Do districts with smaller classes have higher test scores?

FIGURE 4.2 **Scatterplot of Test Score vs. Student-Teacher Ratio (California School District Data)**

Data from 420 California school districts. There is a weak negative relationship between the student-teacher ratio and test scores: the sample correlation is –0.23.



Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

Numerical evidence on whether districts with low STRs have higher test scores

- Compare average test scores in districts with low STRs to those with high STRs ("estimation")
- Test the hypothesis that the mean test scores in the two types of districts are the same, against the alternative hypothesis that they differ ("hypothesis testing")
- Estimate an interval for the difference in the mean test scores, high v. low STR districts ("confidence interval")

Compare districts with “small” (STR $>$ 20) and “large” (STR $\geq$ 20) class sizes

Class Size	Average score ($\bar{y}$)	Standard deviation ($\bar{s_y}$)	n
Small	657.4	19.4	238
Large	650.0	17.9	182

- Estimation of Δ= difference between group means
- Test the hypothesis $\Delta = 0$
- Construct a confidence interval for Δ

Estimation

$$\bar{Y}_{small} - \bar{Y}_{large} = 657.4 - 650.0 = 7.4$$

- $\bar{Y}_{small} = \frac{1}{n_{small}} \sum_{i=1}^{n_{small}} Y_i$
- $\bar{Y}_{large} = \frac{1}{n_{large}} \sum_{i=1}^{n_{large}} Y_i$
- How big is the stdev across districts? 19.1
- What is the diff between 60th and 75th percentile of test score distribution: 667-6559.4=8.2
- Is that a big difference? In practical terms yes (parents and school administration should worry about this!)

Hypothesis testing

Difference-in-means test: compute the t-statistic

$$t = \frac{\bar{Y}_s - \bar{Y}_l}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}} = \frac{\bar{Y}_s - \bar{Y}_l}{SE(\bar{Y}_s - \bar{Y}_l)}$$

- $SE(\bar{Y}_s - \bar{Y}_l)$ is the "standard error" of $\bar{Y}_s - \bar{Y}_l$
- subscripts s and l refer to "small" and "large" STR districts, respectively
- $s_s^2 = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (Y_i - \bar{Y}_s)^2$
- $s_l^2 = \frac{1}{n_l - 1} \sum_{i=1}^{n_l} (Y_i - \bar{Y}_l)^2$

Compute the difference-of-means t-statistic

Size	$(\bar{Y})$	$(s_{\bar{Y}})$	n
small	657.4	19.4	238
large	650.0	17.9	182

- $t = \frac{\bar{Y}_s - \bar{Y}_l}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}} = \frac{657.4 - 650.0}{\sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}}} = \frac{7.4}{1.83} = 4.05$
- $|t| > 1.96$ Reject (at 5% significance level) then null hypothesis that the two mean are the same (equal)

Confidence interval

A 95% confidence interval for the difference between the means is

- $(\bar{Y}_s - \bar{Y}_l) \pm 1.96 \times SE(\bar{Y}_s - \bar{Y}_l) = 7.4 \pm 1.96 \times 1.83 = (3.8, 11.0)$
- Two equivalent statements:
  - The 95% confidence interval for Δ doesn't include 0
  - The hypothesis that $\Delta = 0$ is rejected at the 5% level

Review of statistics

- What is the underlying framework (statistical inference)?
- Estimation: Why estimate $\Delta = (\bar{Y}_s - \bar{Y}_l)$?
- Testing: Why reject $\Delta = 0$ if $|t| > 1.96$?
- Confidence Intervals: What is a confidence interval?

The class size/test score policy question

- What is the effect on test scores of reducing STR by one student/class?
- Policy interest: $\frac{\Delta \text{Test score}}{\Delta STR}$
- Slope of the line relating test score and STR

Population regression line

$$\text{Test Score} = \beta_0 + \beta_1 STR$$

- $\beta_1 = \frac{\Delta \text{Test score}}{\Delta STR}$: slope of population regression line
- β_0 and β_1 are population parameters
- Since we don't know β_1 we must estimate it using data
- Use the least squares ("Ordinary Least Squares" or "OLS") estimator of the unknown parameters β_0 and β_1
- The OLS estimator minimizes the average squared difference between the actual values of Y_i and the prediction (predicted value) based on the estimated line
- Solving the minimization problem yields the **OLS estimators** of β_0 and β_1

Why use OLS, rather than some other estimator?

- OLS is a generalization of the sample average: if the "line" is just an intercept (no X), then the OLS estimator is just the sample average of $Y_1, ..., Y_n$, i.e., $\bar{Y}$
- Like $\bar{Y}$ the OLS estimator has some desirable properties: under certain assumptions, it is unbiased i.e., $E(\hat{\beta}_1) = \beta_1$

OLS Estimator, Predicted Values and Residuals

The OLS estimators of the slople β_1 and the intercept β_0 are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{S_{XY}}{S_X^2} \tag{1}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{2}$$

The OLS predicted values $\hat{Y}_i$ and the residuals $\hat{u}_i$ are:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, ..., n \tag{3}$$

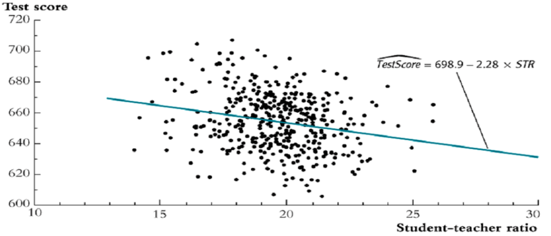
$$\hat{u}_i = Y_i - \hat{Y}_i, \quad i = 1, ..., n \tag{4}$$

The estimated intercept ($\hat{\beta}_0$), slope ($\hat{\beta}_1$) and residual ($\hat{u}_i$) are computed from a sample of n observations of X_i and Y_i, where $i = 1, ..., n$. These are estimates of the unknown true population intercept (β_0), slope (β_1) and error term (u_i).

Application to the California Test Score (TS)-Class Size data (STR)

FIGURE 4.3 **The Estimated Regression Line for the California Data**

The estimated regression line shows a negative relationship between test scores and the student-teacher ratio. If class sizes fall by 1 student, the estimated regression predicts that test scores will increase by 2.28 points.



Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

- Estimated slope $\hat{\beta}_1 = -2.28$
- Estimated intercept $\hat{\beta}_0 = 698.9$
- Estimated regression line $\hat{\text{TS}} = 698.9 - 2.28STR$

Interpretation

$$\hat{TS} = 698.9 - 2.28STR$$

- Districts with one more student per teacher on average have test scores that are 2.28 points lower
- The intercept means that, districts with zero students per teacher would have a (predicted) test score of 698.9
- This interpretation of the intercept makes no sense - it extrapolates the line outside the range of the data - in this application, the intercept is not itself economically meaningful

Predicted values & residuals

$$\hat{TS} = 698.9 - 2.28STR$$

- One of the districts in the data set is Antelope, CA, for which $STR = 19.33$ and $TS = 657.8$
- predicted value: $\hat{Y}_{\text{Antelope}} = 698.6 - 2.28 \times 19.33 = 654.8$
- residual: $\hat{u}_{\text{Antelope}} = 657.8 - 654.8 = 3.0$

OLS regression: STATA output

```
regress testscr str, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  1,   418) =   19.26
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0512
                                                       Root MSE      =  18.581

-------------------------------------------------------------------------
        |               Robust
testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+----------------------------------------------------------------
    str |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
  _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
-------------------------------------------------------------------------
```

$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$

Review OLS

- The OLS regression line is an estimate, computed using our sample of data; a different sample would have given a different value of $\hat{\beta}_1$
- How can we:
  - quantify the sampling uncertainty associated with $\hat{\beta}_1$?
  - use $\hat{\beta}_1$ to test the hypothesis $\beta_1 = 0$?
  - construct a confidence interval for $\beta_1 = 0$?
- Like estimation of the mean
  - The probability framework for linear regression
  - Estimation
  - Hypothesis Testing
  - Confidence intervals

OLS estimate of the TS/STR relation

$$\hat{TS} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28}\, STR, \; R^2 = .05, \; SER = 18.6$$

Is this a credible estimate of the causal effect on test scores of a change in the student-teacher ratio?

- No!
- There are omitted confounding factors (family income; whether the students are native English speakers) that bias the OLS estimator: STR could be "picking up" the effect of these confounding factors
- The bias in the OLS estimator that occurs as a result of an omitted factor is called **omitted variable bias**
- Include English Language Ability (EL) as additional covariate

Additional Covariates: Review Multiple Regression

TABLE 5.1 Differences in Test Scores for California School Districts with Low and High Student Teacher Ratios, by the Percentage of English Learners in the District

	Student-Teacher Ratio < 20		Student-Teacher Ratio ≥ 20		Difference in Test Scores, Low vs. High STR	
	Average Test Score	*n*	Average Test Score	*n*	Difference	*t*-statistic
All Districts	657.4	238	650.0	182	7.4	4.04
Percent of English Learners						
< 2.2%	664.1	78	665.4	27	−1.3	−0.44
2.2–8.8%	666.1	61	661.8	44	4.3	1.44
8.8–23.0%	654.6	55	649.7	50	4.9	1.64
> 23.0%	636.7	44	634.8	61	1.9	0.68

Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

1. Districts with fewer English Learners have higher test scores
2. Districts with lower percent EL (PctEL) have smaller classes
3. Among districts with comparable PctEL the effect of class size is small (recall overall "test score gap" $= 7.4$)

Additional Covariates: Review Multiple Regression

Multiple regression in STATA

```
reg testscr str pctel, robust;

Regression with robust standard errors                 Number of obs =     420
                                                       F(  2,   417) =  223.82
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4264
                                                       Root MSE      =  14.464

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -1.101296   .4328472    -2.54   0.011    -1.95213   -.2504616
       pctel |  -.6497768   .0310318   -20.94   0.000    -.710775   -.5887786
       _cons |   686.0322   8.728224    78.60   0.000    668.8754     703.189
------------------------------------------------------------------------------
```

$$\widehat{TestScore} = 696.0 - 1.10 \times STR - 0.65 PctEL$$

Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

Non-linear relations between dependent and independent vars

1. The approximation that the regression function is linear might be good for some variables, but not for others.
2. The multiple regression framework can be extended to handle regression functions that are nonlinear in one or more Xs
3. e.g., the Test Score – average district income relation

Linear and Non-linear relatioships

FIGURE 4.3 The Estimated Regression Line for the California Data

The estimated regression line shows a negative relationship between test scores and the student-teacher ratio. If class sizes fall by 1 student, the estimated regression predicts that test scores will increase by 2.28 points.

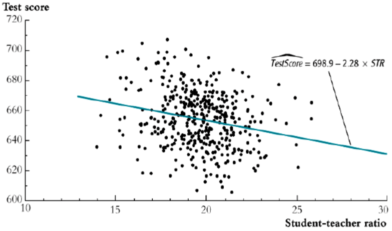
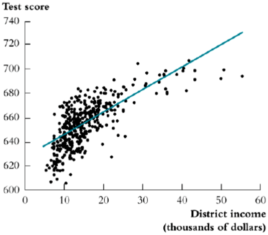


FIGURE 6.2 Scatterplot of Test Score vs. District Income with a Linear OLS Regression Function

There is a positive correlation between test scores and district income (correlation = 0.71), but the linear OLS regression line does not adequately describe the relationship between these variables.



Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

Non-linear relations between dependent and independent vars

If a relation between Y and X is nonlinear:

1. the effect on Y of a change in X depends on the value of X - that is, the marginal effect of X is not constant
2. the linear regression is misspecified - the functional form is wrong
3. the estimator of the effect on Y of X is biased - it needn't even be right on average
4. the solution to this is to estimate a regression function that is nonlinear in X

The General Nonlinear Population Regression Function

$$Y_i = f(X_{1i}, X_{2i}, X_{3i}, .., X_{ki}) + u_i,\ i = 1, ..., n$$

- Assumptions
  1. $E(u_i|X_{1i}, X_{2i}, X_{3i}, .., X_{ki}) = 0$: f is the conditional expectation of Y given Xs
  2. $(X_{1i}, X_{2i}, X_{3i}, .., X_{ki}, Y_i)$ are i.i.d
  3. "enough" moments exist but depend on specific f
  4. no perfect multicollinearity: depend on specific f

Non-linear relatioships: The expected effect on Y of a change in a specific X

The expected change in Y (i.e., ΔY) associated with the change in X_1 (i.e., ΔX_1) holding $X_2, ..., X_k$ constant, is the difference between the value of the population regression function before and after changing X_1, holding $X_2, ..., X_k$ constant. That is, the expected change in Y is the difference:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, ..., X_k) - f(X_1, X_2, ..., X_k) \qquad (5)$$

The estimator of this unknown population difference is the difference between the predicted values of these two cases if we assume that $f(X_1, X_2, ..., X_k)$ is the predicted values of Y based on the estimator $\hat{f}$ of the population regression function. Then the predicted change in Y is:

$$\Delta \hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, ..., X_k) - \hat{f}(X_1, X_2, ..., X_k) \qquad (6)$$

Two complementary approaches

- Polynomials in X
  - The population regression function is approximated by a quadratic, cubic, or higher-degree polynomial
- Logarithmic transformations
  - Y and/or X is transformed by taking its logarithm
  - this gives a "percentages" interpretation that makes sense in many applications

Polynomials in X

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

- This is just the linear multiple regression model – except that the regressors are powers of X
- Estimation, hypothesis testing, etc. proceeds as in the multiple regression model using OLS
- The coefficients are difficult to interpret, but the regression function itself is interpretable

Example: the TestScore – Income relation

- Income_i= average district income in the ith district (in thousdand dollars per capita)
- Quadratic specification:
  - $\text{Income}_i = \beta_0 + \beta_1 \text{Income}_i + \beta_2 (\text{Income}_i)^2 + u_i$
- Cubic specification:
  - $\text{Income}_i = \beta_0 + \beta_1 \text{Income}_i + \beta_2 (\text{Income}_i)^2 + \beta_3 (\text{Income}_i)^3 + u_i$

Non-linear relatioships

Estimation of the quadratic specification in STATA

```
generate avginc2 = avginc*avginc;        Create a new regressor
reg testscr avginc avginc2, r;

Regression with robust standard errors                 Number of obs =     420
                                                       F(  2,   417) =  428.52
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5562
                                                       Root MSE      =  12.724

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      avginc |   3.850995   .2680941    14.36   0.000      3.32401    4.377979
     avginc2 |  -.0423085   .0047803    -8.85   0.000    -.051705   -.0329119
       _cons |   607.3017   2.901754   209.29   0.000     601.5978    613.0056
------------------------------------------------------------------------------
```

The t-statistic on $Income^2$ is -8.85, so the hypothesis of linearity is rejected against the quadratic alternative at the 1% significance level.

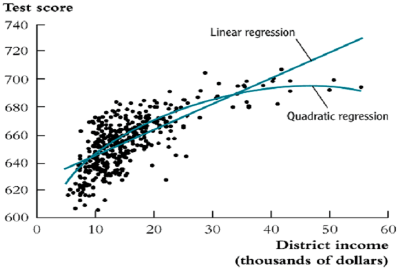
Non-linear relatioships: estimated regression function

Plot predicted values

$$\hat{TS} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85}\,\text{Income}_i - \underset{(0.005)}{0.042}\text{Income}_i^2$$

FIGURE 6.3 **Scatterplot of Test Score vs. District Income with Linear and Quadratic Regression Functions**

The quadratic OLS regression function fits the data better than the linear OLS regression function.



Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

Non-linear relatioships: Review polynomial regression functions

Compute "effects" for different values of X

- Predicted change in TS for a change in income to \$6,000 from \$5,000 per capita:
  - $\Delta\hat{TS} =$
    $607.3+3.85\times6-0.042\times6^2-(607.3+3.85\times5-0.042\times5^2) = 3.4$
- Predicted change in TS for a change in income to \$26,000 from \$25,000 per capita:
  - $\Delta\hat{TS} = 1.7$
- Predicted change in TS for a change in income to \$46,000 from \$45,000 per capita:
  - $\Delta\hat{TS} = 0.0$
- The "effect" of a change in income is greater at low than high income levels (perhaps, a declining marginal benefit of an increase in school budgets?) Caution! What about a change from 65 to 66? Don't extrapolate outside the range of the data

Logarithmic functions of Y and/or X

- $ln(X)$: natural logarithm of X
- Logarithmic transforms permit modeling relations in "percentage" terms (like elasticities), rather than linearly.
- Why?
  - $ln(x + \Delta x) - ln(x) = ln(1 + \frac{\Delta x}{x}) \equiv \frac{\Delta x}{x}$, calculus $\frac{d ln(x)}{dx} = \frac{1}{x}$
- Numerically
  - $ln(1.01) = .00995 \equiv .01$, $ln(1.10) = .0953 \equiv .10$

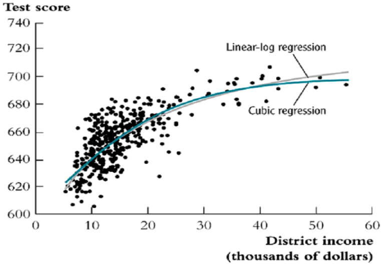
Non-linear relatioships: estimated regression function

Plot predicted values: Logarithmic transformation

$$\hat{TS} = 557.8 + 36.42 \times ln\text{Income}_i$$

FIGURE 6.7 **The Linear-Log and Cubic Regression Functions**

The estimated cubic regression function (Equation (6.11)) and the estimated linear-log regression function (Equation (6.18)) are nearly identical in this sample.



Source: J.H. Stock and M.W. Watson, Introduction to Econometrics (first edition), Addison-Wesley, 2003

Logarithmic transformations of variables

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Logarithmic transformations: four possible combinations
  - Linear (no transformations)
  - Linear-Log model
  - Log-Linear model
  - Log-Log model

Logarithmic transformations of variables

Figure 1: Combinations of logarithmic transformations

	Variable X	
Variable Y	X	$logX$
Y	Linear	Linear-Log
Estimated model	$\hat{Y}_i = \beta_0 + \beta_1 X_i$	$\hat{Y}_i = \beta_0 + \beta_1 logX_i$
$logY$	Log-Linear	Log-Log
Estimated model	$log\hat{Y}_i = \beta_0 + \beta_1 X_i$	$log\hat{Y}_i = \beta_0 + \beta_1 logX_i$

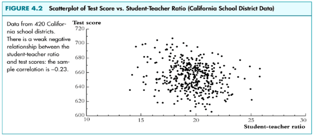
Logarithmic transformations of variables

Review: Properties of logarithms and exponential functions

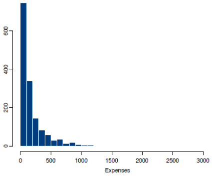
- $log(e) = 1$
- $log(1) = 0$
- $log(x^A) = Alog(x)$
- $log(e)^A = A$
- $e^{log(A)} = A$
- $log(A \times B) = log(A) + log(B)$
- $log(\frac{A}{B}) = log(A) - log(B)$
- $e^{A \times B} = (e^A)^B$
- $e^{A+B} = (e^A) \times (e^B)$
- $e^{A-B} = \frac{e^A}{e^B}$

Logarithmic transformations of variables

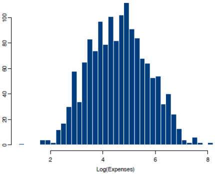
- capture non-linear relationship between the independent and dependent variables (e.g., $Y_i = \beta_0 + \beta_1 log X_i + u_i$)



- transform a highly skewed variable into an approximately normal variable



(a) skewed



(b) normal

Logarithmic transformations of variables

Interpretation: Linear model

- $Y_i = \beta_0 + \beta_1 X_i + u_i$
- $\hat{\beta}_1$
- Change in Y for a one-unit change in X

Logarithmic transformations of variables

Interpretation: Linear-Log model

- $Y_i = \beta_0 + \beta_1 logX_i + u_i$
- $\hat{\beta}_1$
- A one-unit increase in *logX* will produce an expected increase in Y of $\hat{\beta}_1$ units.
- Example
  - $\hat{Y}_i = 450.2 + 65.32 logX_i$, where Y is the average math SAT score and X is the expenditure per student ($i = 1, ..., N$ schools).
  - $\hat{\beta}_1 = 65.32$: a 1 percent increase in expenditure per student increases the average math SAT score by 0.65 points (i.e., $\hat{\beta}_1/100$ or $65.32/100$).

Logarithmic transformations of variables

Interpretation: Log-Linear model

- $logY_i = \beta_0 + \beta_1 X_i + u_i$
- $\hat{\beta}_1$
- A one-unit increase in X will produce an expected increase in $logY$ of $\hat{\beta}_1$ units.
- Example
  - $\hat{Y}_i = 10.5 + 0.08 logX_i$, where Y is the annual earnings and X is the years of completed schooling per worker ($i = 1, ..., N$ workers).
  - $\hat{\beta}_1 = 0.08$: a 1 unit increase in years of schooling (1 more year) increases annual earnings by 8% (i.e., $\hat{\beta}_1 \times 100$ or 0.08×100).

Logarithmic transformations of variables

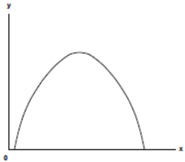
Interpretation: Log-Log model

- $logY_i = \beta_0 + \beta_1 logX_i + u_i$
- $\hat{\beta}_1$
- Expected percentage change in Y when X increases by some percentage (e.g., 1% or 10%). Directly estimate elasticities.
- Example
  - $\hat{Y}_i = 7.09 - 0.50 logX_i$, where Y is the percentage of urban population and X is the per capita GDP per country ($i = 1, ..., N$ countries).
  - 1% increase in X: $\hat{\beta}_1 = 0.50$: a 1% increase GDP reduces urban population by 0.5% (i.e., $\hat{\beta}_1 \times 1$ or 0.50×1).
  - 10% increase in X: $\hat{\beta}_1 = 0.50$: a 10% increase GDP reduces urban population by 5.0% (i.e., $\hat{\beta}_1 \times 10$ or 0.50×10).

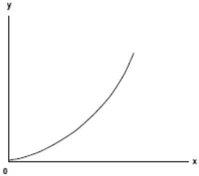
Higher order polynomials

Quadratic transformation

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



(c) $\beta_1 > 0$ & $\beta_2 < 0$

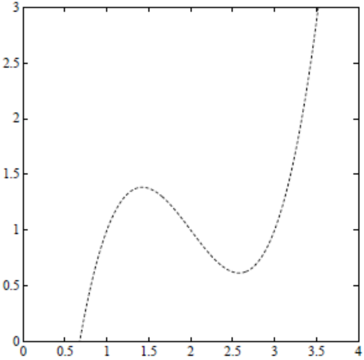


(d) $\beta_1 > 0$ & $\beta_2 > 0$

Higher order polynomials

Cubic transformation

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$



Interaction effects

Model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + u_i$$

- "main terms": X_{1i} and X_{2i}
- "interaction terms": $X_{1i} X_{2i}$
- partial derivative of Y wrt X_1: $\beta_1 + \beta_3 X_{2i}$
- if $X_{2i} = 0$ then Y depends on X
- test $\beta_1 = 0$: no effect of X_1 on Y when $X_2 = 0$

Interaction effects

In models with multiplicative terms, the regression coefficients for X_1 and X_2 reflect *conditional* relationships. β_1 of the effect of X_1 on Y when $X_2 = 0$. Similarly, β_2 is the effect of X_2 on Y when $X_1 = 0$. For example, we get

$$\begin{aligned} Y &= \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 X_2) + \epsilon \\ &= \alpha + \beta_1 X_1 + \beta_2 0 + \beta_3 (X_1 0) + \epsilon \\ &= \alpha + \beta_1 X_1 + \epsilon \end{aligned}$$

So, we can say that, for a person with $X_2 = 0$, a 1 unit increase in X_1 will produce, on average, a β_1 unit increase in Y.

Interaction effects

However, suppose that someone has a score of 3 on X_2. The effect X_1 is then

$$\begin{aligned} Y &= \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 X_2) + \epsilon \\ &= \alpha + \beta_1 X_1 + \beta_2 3 + \beta_3 (X_1 3) + \epsilon \\ &= \alpha + \beta_1 X_1 + 3\beta_2 + 3\beta_3 X_1 + \epsilon \\ &= \alpha + 3\beta_2 + (\beta_1 + 3\beta_3) X_1 + \epsilon \end{aligned}$$

So, when $X_2 = 3$, a 1 unit increase in X_1 will produce, on average, a $\beta_1 + 3\beta_3$ unit increase in Y.

Interaction effects: Review

1. Perhaps a class size reduction is more effective in some circumstances than in other...
2. Perhaps smaller classes help more if there are many English learners, who need individual attention
3. How to model such "interactions" between X_1 and X_2?
4. Continuous and/or Binary Vars