

# 第三讲：多元线性回归和 Stata 操作

于明哲

北京工商大学 国际经管学院

2023 暑期课堂

# 本章内容

- 1 回顾：一元线性回归模型
- 2 多元回归：动机
- 3 多元线性回归模型
- 4 假设检验
- 5 多元回归的模型设定
- 6 实证分析的基本步骤

# 当前章节

① 回顾：一元线性回归模型

② 多元回归：动机

③ 多元线性回归模型

④ 假设检验

⑤ 多元回归的模型设定

⑥ 实证分析的基本步骤

## 例子：小班教育与教学成果 I

- 小学教育改革的理念：提高小学教育质量的方法之一是减少班级中的人数
  - 老师可以更多地关注每一名学生
  - 课堂纪律更好
- 然而，小班教育的成本很高
  - 需要雇佣更多的老师
  - 建造更多的教室
- 决策者需要根据准确的数据分析结果，来权衡（Tradeoff）小班教育的收益和成本

## 例子：小班教育与教学成果 II

- 数据分析的结果，又叫做“经验证据 (Empirical Evidence)”，即基于数据的证据
- **数据**：1999 年 California 的 420 个学区中的班级规模和学习成绩
- **发现**：在小班授课学区上学的孩子，平均而言学习成绩更好
- **问题**：这是否意味着小班授课有益于学习成绩？
  - 小班授课的学区通常居民更加富有，其子女有更多的课外学习机会和更多的见识，
  - 更好的学习成绩可能来自“额外地”学习机会，而不是课堂的规模
- 从多元回归开始，我们将讨论如何控制其他因素的影响

# 其他经济学中问题

- 劳动市场或贷款市场存在性别歧视吗？（劳动经济学）
- 增加健康支出会改善居民的健康状况吗？（公共经济学）
- 企业规模会对市场效率产生影响吗？（产业组织理论）
- 网络的普及会提高落后地区居民的受教育水平吗？（发展经济学）
- 制度差异会导致经济表现差异吗？（经济增长）
- ...

# 因果效应与理想化实验

- 上页中的问题涉及当代计量经济学的重要任务之一：因果推断 (Causal Inference)
- **Causality**: a specific action (applying fertilizer) leads to a specific, measurable consequence (more tomatoes).
  - Action: applying fertilizer
  - Consequence: more tomatoes

# 因果效应与理想化实验

- 如何估计因果效应：理想化实验
- 特别注意：暂没有其他方法可以估计因果效应
- 理想化实验：随机化控制实验 (Randomized controlled experiment)
  - 控制组 (Control group)：不接受处理
  - 处理组 (Treatment group)：接受处理

## Causal Effect

The effect on an outcome of a given treatment, as measured in an ideal randomized controlled experiment.



# 因果效应与理想化实验

- 特别注意：有效的随机化控制实验必须要保证处理组与控制组的结果差异仅由实验中施加的处理所导致
- 有了理想化实验的理念，就可以回答之前提出的几个经济学问题
- 出于现实原因 (unethical, impossible to execute satisfactorily, too time-consuming, or prohibitively expensive)，社会科学研究通常无法开展随机化控制实验
- 当代计量经济学：依靠非实验性数据，通过方法 (Method)，尽量模拟出随机化控制实验的结果

# 数据的来源 I

- 实验数据 (Experimental Data):
  - 通常来自评估 Treatment 或 Policy 的效果的实验，获取数据的目的是为了估计因果效应
- 然而，由于获取实验数据存在种种问题（伦理的、成本的，操作的），因此经济学中更大一部分研究数据都来自现实经济活动的观测数据
- 观测数据 (Observational Data):
  - 来自调查 (Survey)，或 Administrative Records（比如个人的税收记录，贷款的申请记录，淘宝、京东、滴滴、美团、抖音……）

## 数据的来源 II

- 由于观测数据并不是产生于随机化实验，因此直接使用观测数据估计因果关系往往存在严重的问题
- 从某种意义上说，当代计量经济学的主要任务之一，就是找到使用观测数据估计因果关系的方法
- 本次课程的主要目的，就是学习使用观测数据估计因果关系的几个基础方法
- 虽然 2023 年秋季学期的《计量经济学》也是 Introduction Level 的课程，但会包含更多细节

# 数据的类型

- 无论是实验数据还是观测数据，基本上只有以下三种数据结构：
  - 横截面数据 (Cross-sectional Data)
  - 时间序列数据 (Time Series Data)
  - 面板数据 (Panel Data)

# 横截面数据 (Cross-sectional Data)

- 同一时间范围内，来自不同主体 (Entities) 的数据
- 例如，加州各学区 (Entity) 的平均考试成绩及其他相关变量

# 横截面数据 (Cross-sectional Data)

**TABLE 1.1** Selected Observations on Test Scores and Other Variables for California School Districts in 1999

Observation (District) Number	District Average Test Score (fifth grade)	Student-Teacher Ratio	Expenditure per Pupil (\$)	Percentage of Students Learning English
1	690.8	17.89	\$6385	0.0%
2	661.2	21.52	5099	4.6
3	643.6	18.70	5502	30.0
4	647.7	17.36	7102	0.0
5	640.8	18.67	5236	13.9
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
418	645.0	21.89	4403	24.3
419	672.2	20.20	4776	3.0
420	655.8	19.04	5993	5.0

*Note:* The California test score data set is described in Appendix 4.1.

# 时间序列数据 (Time Series Data)

- 同一个主体在不同时间点上的各变量取值
- 例如，美国 (Entity) 1960 年以来各个季度的 GDP 增长率以及期限利差

# 时间序列数据 (Time Series Data)

**TABLE 1.2** Selected Observations on the Growth Rate of GDP and the Term Spread in the United States: Quarterly Data, 1960:Q1–2013:Q1

Observation Number	Date (year:quarter)	GDP Growth Rate (% at an annual rate)	Term Spread (% per year)
1	1960:Q1	8.8%	0.6%
2	1960:Q2	−1.5	1.3
3	1960:Q3	1.0	1.5
4	1960:Q4	−4.9	1.6
5	1961:Q1	2.7	1.4
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
211	2012:Q3	2.7	1.5
212	2012:Q4	0.1	1.6
213	2013:Q1	1.1	1.9

*Note:* The United States GDP and term spread data set is described in Appendix 14.1.



# 面板数据 (Panel Data)

- 不同主体在不同时间点上的数据
- 例如，美国不同州 (Entities) 在不同年份的烟草销售数量以及其他变量的数据

# 面板数据 (Panel Data)

**TABLE 1.3** Selected Observations on Cigarette Sales, Prices, and Taxes, by State and Year for U.S. States, 1985–1995

Observation Number	State	Year	Cigarette Sales (packs per capita)	Average Price per Pack (including taxes)	Total Taxes (cigarette excise tax + sales tax)
1	Alabama	1985	116.5	\$1.022	\$0.333
2	Arkansas	1985	128.5	1.015	0.370
3	Arizona	1985	104.5	1.086	0.362
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
47	West Virginia	1985	112.8	1.089	0.382
48	Wyoming	1985	129.4	0.935	0.240
49	Alabama	1986	117.2	1.080	0.334
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
96	Wyoming	1986	127.8	1.007	0.240
97	Alabama	1987	115.8	1.135	0.335
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
528	Wyoming	1995	112.2	1.585	0.360

*Note:* The cigarette consumption data set is described in Appendix 12.1.

# 一元线性回归模型

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (1)$$

- 方程 (1) 称为 “一元线性回归模型” (linear regression model with a single regressor)
- 其中
  - ①  $Y$ : 被解释变量 (dependent variable)
  - ②  $X$ : 解释变量 (independent variable)
  - ③  $\beta_0, \beta_1$ : 回归系数 (Coefficient)
  - ④  $u_i$ : 误差项 (Error Term)

# 一元回归：学生教师比

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

- 当你开展关于班级规模与考试成绩关系的研究时，就可以使用一元回归模型
- 假设你有一组  $n$  个学区组成的样本，并且
  - ①  $Y_i$  表示学区  $i$  的平均考试成绩
  - ②  $X_i$  表示学区  $i$  的平均班级规模
  - ③  $u_i$  表示影响学区  $i$  平均成绩的其他因素，那么就有
- 问题： $\beta_0, \beta_1$  的经济含义是什么？

# 误差项

- $u_i$ : 表示除  $X$  之外的, 所有影响  $Y$  的因素
- 不要以为  $u_i$  只是个小问题, 计量经济学的所有问题几乎都因  $u_i$  而起!
- 在开始学习时, 为了保持模型的简单性, 通常假设  $u$  和  $X$  独立, 即
$$E(u|X) = 0$$
- 因此,

$$E(Y|X) = \beta_0 + \beta_1 X$$

- $\beta_0 + \beta_1 X$  称为“总体回归方程”(Population Regression Function), 即总体平均意义上  $X$  和  $Y$  的关系

# OLS regression: STATA output

$Y$        $X$   
 regress testscr str, robust

Regression with robust standard errors

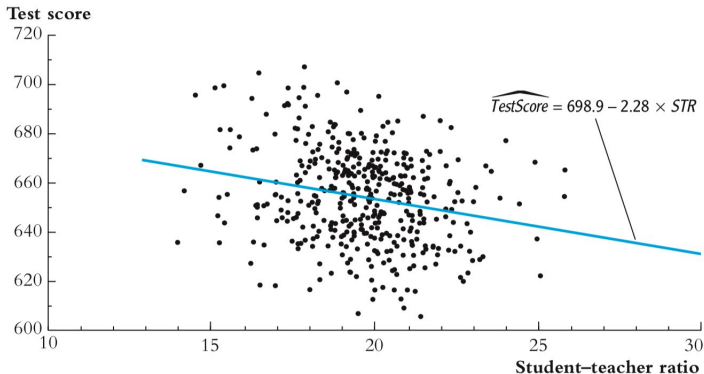
Number of obs = 420  
 F( 1, 418) = 19.26  
 Prob > F = 0.0000  
 R-squared = 0.0512  
 Root MSE = 18.581

-----						
testscr		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
-----						
$\beta_1$ str		-2.279808	.5194892	-4.39	0.000	-3.300945 -1.258671
$\beta_0$ _cons		698.933	10.36436	67.44	0.000	678.5602 719.3057
-----						

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

**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 one student, the estimated regression predicts that test scores will increase by 2.28 points.



# 当前章节

1 回顾：一元线性回归模型

2 多元回归：动机

3 多元线性回归模型

4 假设检验

5 多元回归的模型设定

6 实证分析的基本步骤



# 遗漏变量偏误

- 在班级规模 ( $X$ ) 与考试成绩 ( $Y$ ) 的研究中, 显然影响考试成绩的因素不止有一个
- 从理论上说, 为什么会有误差项  $u$ , 就是因为  $u$  里包含了决定  $Y$  的其他因素
- 也就是说, 回归模型中总是存在遗漏变量 (Omitted Variables)
- 有些时候, 遗漏变量会造成 OLS 估计量存在偏误

- 在班级规模与考试成绩的研究中，可能存在三个遗漏变量
  - ① 考试的时间
  - ② 学校的教室数量（相对于学生数量）
  - ③ 英语学习者的比例（外来移民的比例）
- 问题：遗漏以上哪个变量可能导致偏误？
- 遗漏变量偏误 (Omitted Variable Bias): If the regressor is correlated with a variable that has been omitted from the analysis and that determines, in part, the dependent variable, then the OLS estimator will have omitted variable bias.

# 遗漏变量偏误：数学表达

- 设  $Z$  是回归模型的遗漏变量，如果遗漏变量  $Z$  会导致模型存在遗漏变量偏误，需要同时满足以下两个条件：
  - ①  $Z$  是  $Y$  的决定因素之一 (i.e.  $Z$  是  $u$  的决定因素之一， $\text{corr}(Z, u) \neq 0$ )
  - ②  $Z$  和  $X$  相关 (i.e.  $\text{corr}(Z, X) \neq 0$ )
- 因此，英语学习者的比例会导致遗漏变量偏误

# 班级规模与考试成绩：遗漏变量

**TABLE 6.1** 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
Percentage of English learners						
< 1.9%	664.5	76	665.4	27	-0.9	-0.30
1.9-8.8%	665.2	64	661.8	44	3.3	1.13
8.8-23.0%	654.9	54	649.7	50	5.2	1.72
> 23.0%	636.7	44	634.8	61	1.9	0.68

学区中不同的英语学习者（移民）比例会导致显著的平均成绩差异!!

# 遗漏变量偏误：公式

- 如何计算遗漏变量偏误

$$\begin{aligned}\hat{\beta}_1 - \beta_1 &= \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) u_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \\ &\xrightarrow{p} \frac{\sigma_{Xu}}{\sigma_X^2} \\ &= \left( \frac{\sigma_u}{\sigma_X} \right) \times \left( \frac{\sigma_{Xu}}{\sigma_X \sigma_u} \right) = \left( \frac{\sigma_u}{\sigma_X} \right) \rho_{Xu},\end{aligned}$$

where  $\rho_{Xu} = \text{corr}(X, u)$ .

- 如果假设 1 成立 ( $E(u_i|X_i) = 0$ ) , 那么  $E((X_i - \mu_X)u_i) = \text{cov}(X_i, u_i) = 0$   
(即  $\rho_{Xu} = 0$ ) , 且  $E(\hat{\beta}_1) = \beta_1$

## 遗漏变量偏误：后果

$$\hat{\beta}_1 \xrightarrow{p} \beta_1 + \left( \frac{\sigma_u}{\sigma_X} \right) \rho_{Xu}$$

- 根据上述公式，遗漏变量偏误会造成以下后果
  - 大样本下回归系数估计量仍然存在偏误
  - 偏误的大小取决于  $|\rho_{Xu}|$  的大小
  - 偏误的方向由  $\rho_{Xu}$  的方向决定

- 遗漏变量偏误对研究结果会造成什么影响？
  - 错误判断因果关系！
- Definition of Causal Effect in Econometrics: the effect that is measured in an **Ideal Randomized Controlled Experiment**.
- 解决遗漏变量偏误的方法
  - ① 如果遗漏变量**可观测**:
    - 把遗漏变量作为新的解释变量引入回归方程
  - ② 如果遗漏变量**不可观测**
    - 找到一个随机化可控实验（准实验）

# 当前章节

1 回顾：一元线性回归模型

2 多元回归：动机

3 多元线性回归模型

4 假设检验

5 多元回归的模型设定

6 实证分析的基本步骤



# 线性回归模型：两个解释变量

- 考虑带有两个解释变量的回归模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad i = 1, \dots, n$$

- $(Y_i, X_{1i}, X_{2i})$  是关于  $(Y, X_1, X_2)$  的第  $i$  个观测值
- $\beta_0$ : 未知的截距项
- $\beta_1$ : 其他条件不变时,  $X_1$  的变化对  $Y$  的影响
- $\beta_2$ : 其他条件不变时,  $X_2$  的变化对  $Y$  的影响
- $u_i$ : 误差项 (遗漏变量)

# 回归系数的含义

- 对于模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad i = 1, \dots, n$$

- 什么叫“ $X_1$  的变化对  $Y$  的影响”
- 给  $X_1$  一个变化量  $\Delta X_1$ ，同时保持  $X_2$  不变，那么在变化以前，回归方程为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i$$

- $X_1$  改变后，回归方程变为

$$Y + \Delta Y = \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2 + u_i$$

- 改变之前:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i$
- 改变之后:  $Y + \Delta Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2 + u_i$
- 二者之差:  $\Delta Y = \beta_1 \Delta X_1$

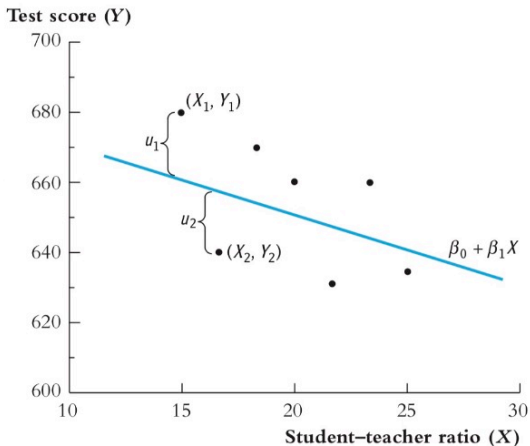
$$\beta_1 = \frac{\Delta Y}{\Delta X_1}, \text{ holding } X_2 \text{ constant}$$

$$\beta_2 = \frac{\Delta Y}{\Delta X_2}, \text{ holding } X_1 \text{ constant}$$

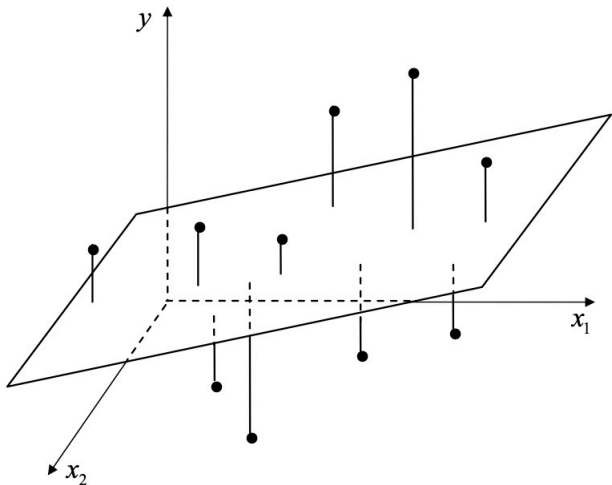
$$\beta_0 = \text{predicted value of } Y \text{ when } X_1 = X_2 = 0.$$

# 多元线性回归模型的系数估计：最小二乘法

- OLS: Ordinary Least Squares
- BLUE: Best Linear Unbiased Estimator



# 多元线性回归模型的系数估计：最小二乘法



# 多元线性回归模型的系数估计：最小二乘法

- 两个解释变量的最小二乘法

$$\min_{b_0, b_1, b_2} \sum_{i=1}^n [Y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i})]^2$$

- 很容易推广到  $k$  个解释变量的模型

$$\sum_{i=1}^n (Y_i - b_0 - b_1 X_{1i} - \cdots - b_k X_{ki})^2.$$

# 班级规模与考试成绩：加入控制变量

## 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

testscr	Coef.	Robust 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

$$\widehat{TestScore} = 686.0 - 1.10 \times STR - 0.65 PctEL$$

## 拟合优度：多元线性回归

- $R^2$  和  $\bar{R}^2$
- $\bar{R}^2$ : Adjusted  $R^2$
- 在多元回归中加入新的解释变量后，只要这个解释变量的回归系数不为零，那么 SSR 一定减小 ( $R^2$ )，所以，我们得调整一下这种影响

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS},$$
$$\bar{R}^2 = 1 - \left( \frac{n-1}{n-k-1} \right) \frac{SSR}{TSS}$$

- 显然， $\bar{R}^2 < R^2$ ，不过当  $n$  较大时，二者的差异会很小
- 事实上，在实证研究中，二者的差异不必太在意！



# 回归系数的估计值

- 我们一直没有说回归系数的估计值是多少，因为有点难！
- 估计系数的方法和之前相同，即求偏导数并通过一阶条件求解系数的估计量
- 但如果这么求解的话，过程会非常复杂
- 通常情况下，使用矩阵描述计量模型会非常简便!!
- 虽然计量软件会帮我们计算出所有结果（统计软件都是使用矩阵进行计算），但如果你想获得 M.A. 或 Ph.D 学位，就需要了解矩阵表达计量模型的方法
- So, let's miss linear algebra

# 多元线性回归模型的假设

- 给定多元线性回归模型

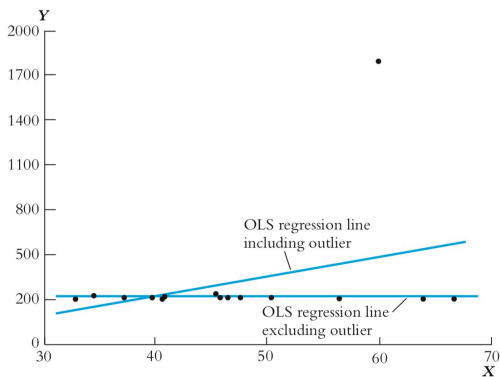
$$Y = X\beta + U$$

- 为了估计  $\beta$ ，需要对多元线性回归模型做出如下假设
  - 1  $E(u_i|\mathbf{X}_i) = 0$  或  $E(\mathbf{U}|\mathbf{X}) = 0$ :  $u_i$  的条件均值为 0
  - 2  $(\mathbf{X}_i, Y_i), i = 1, \dots, n$  独立同分布
  - 3  $0 < E(\mathbf{X}_i) < \infty$  和  $0 < E(u_i) < \infty$  (非零且有限的四阶矩)
  - 4  $\mathbf{X}'\mathbf{X}$  是非奇异的 (Nonsingular)

# Assumption 3: No Large Outliers

**FIGURE 4.5** The Sensitivity of OLS to Large Outliers

This hypothetical data set has one outlier. The OLS regression line estimated with the outlier shows a strong positive relationship between  $X$  and  $Y$ , but the OLS regression line estimated without the outlier shows no relationship.



## Assumption 4: $\mathbf{X}'\mathbf{X}$ is Nonsingular

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

- 如果  $\mathbf{X}'\mathbf{X}$  不是非奇异的 (Nonsingular), 即无法求逆矩阵
- 因此也就无法估计  $\hat{\beta}$
- 如果解释变量  $X_1$ 、 $X_2$  存在线性相关, 那么统计软件会直接帮你 drop 掉其中一个变量

**Example:** Suppose you accidentally include *STR* twice:

```
regress testscr str 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
str	(dropped)					
_cons	698.933	10.36436	67.44	0.000	678.5602	719.3057

# 当前章节

1 回顾：一元线性回归模型

2 多元回归：动机

3 多元线性回归模型

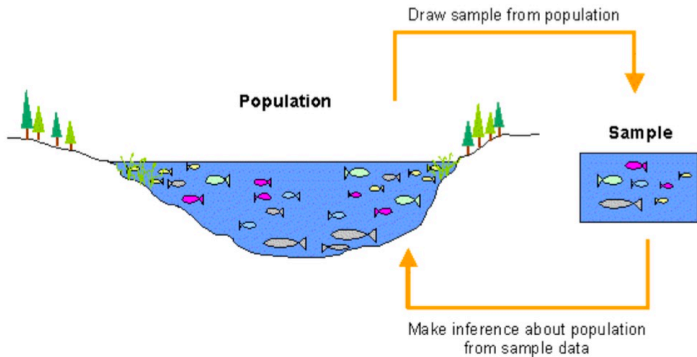
4 假设检验

5 多元回归的模型设定

6 实证分析的基本步骤

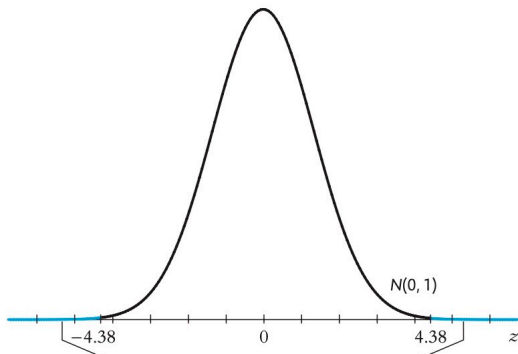
# 单系数的假设检验

- 与一元线性回归模型没有区别！
- 总体 (Population) v.s. 样本 (Sample)



# 单系数的假设检验

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i$
- 原假设和双边备择假设  $H_0 : \beta_1 = 0$  v.s.  $H_1 : \beta_1 \neq 0$





# 单系数的假设检验

## 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

testscr	Coef.	Robust 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

$$\widehat{TestScore} = 686.0 - 1.10 \times STR - 0.65 PctEL$$

- 此时，班级规模（str）变量的  $p$  值为 0.011，可以在 5% 的显著性水平下拒绝原假设

# 当前章节

1 回顾：一元线性回归模型

2 多元回归：动机

3 多元线性回归模型

4 假设检验

5 多元回归的模型设定

6 实证分析的基本步骤

# 如何选择模型中的变量

- 模型设定问题无论是对于初学者还是资深专家，都是主要的挑战之一
- 如何上手：从[遗漏变量偏误的来源](#)开始思考，要根据经济理论找到合适的变量，并试图获得无偏的估计量
- 注意!! 千万不要仅仅依赖某些统计量（比如  $R^2$  或  $p - Value$ ）对模型的质量进行判断

# 遗漏变量偏误

- 如果某个变量是原回归模型的遗漏变量，那么至少满足以下的条件之一
  - ❶ 该变量必须是被解释变量的决定因素
  - ❷ 该变量与至少一个原模型中的解释变量相关
- 例如，在有关 Educational Attainment 的研究中，家庭背景通常是考试成绩的决定因素之一，因为富裕家庭的孩子可以上课外班，并因此提高考试成绩。同时，富人区的学校通常投入更高，这样会导致更低的学生教师比。
- 所以，家庭背景就是之前班级规模与考试成绩模型的遗漏变量
- 存在遗漏变量的后果：回归系数是有偏的

# 模型设定的基本原则

- 解决方法：在回归模型中加入控制变量
- 在我们的例子中，家庭背景通常是不可观测的，但可以找到代理变量（Proxy）
  - 英语学习者的比例
  - 享受午餐优惠的学生比例
  - 享受收入补贴的学生比例

# 当前章节

1 回顾：一元线性回归模型

2 多元回归：动机

3 多元线性回归模型

4 假设检验

5 多元回归的模型设定

6 实证分析的基本步骤

# 写作秘籍：PEEEL

- ❶ P: point, 观点陈述句
- ❷ E: explanation, 解释观点, step by step 一步一步地阐述你的观点, 形成一个 chains of analysis
- ❸ E: example 举例、实证分析
  - 介绍数据和变量
  - 构建模型
  - 得到回归结果并分析
- ❹ E: evaluation 正或反
- ❺ L: link 总结, 首尾呼应