第 18 章　随机实验与自然实验

18.1　实 验 数 据

假设研究x_1是否导致 y。假定$\{x_1, x_2, \cdots, x_K\}$包含所有影响 y 的因素。

不同学科采用不同的实验方法，大致分为以下几类。

(1) 控制实验(controlled experiment)：在理想的物理实验中，控制$\{x_2, \cdots, x_K\}$全部不变，单独让x_1变化，观察 y 的变化。

(2) 随机(控制)实验(randomized controlled experiment)：

【例】医学上对新药x_1疗效的实验。由于参加实验者的体质与生活方式不同，不可能完全控制所有其他因素$\{x_2, \cdots, x_K\}$。

随机实验将实验人群(或个体)随机地分为两组，其中“实验组”或“处理组”(treatment group)服用真药，而“控制组”(control group，也称“对照组”)服用“安慰药”(placebo)。

被试者不知道自己分在哪一组，避免心理干扰。有时科研人员也不知道被试者在哪一组，称为“双盲法”(double blind)。

【例】农学中将地块随机地分成三组(很难找到土壤条件完全一样的地块)，分别给予不同的施肥量，然后考察施肥的效果。

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	2.79	0.014
No Hospital	90049	2.07	0.003

(3) 自然实验或准实验(natural experiment or quasi experiment)：

由于某些并非为了实验目的而发生的外部突发事件，使得当事人仿佛被随机地分在了实验组或控制组。

【例】一个州通过某法律，但相邻州未通过此法律。两州民众事先不知道哪个州会通过此法律，故无法自我选择住在哪个州。从考察法律的效果而言，可近似认为民众随机选择住在哪个州，或被随机分为实验组(通过法律)与控制组(没通过法律)。

(4) 思想实验(thought experiment)：

Milton Friedman 曾设想在小岛上通过空投货币，考察该岛的宏观经济的变化。

18.2　理想的随机实验

在理想的随机实验(ideal randomized experiment)中，实验组与控制组的成员决定完全随机，比如，通过抛硬币或电脑随机数来决定。

个体究竟分在哪一组或得到多大的实验“处理水平”(treatment level)，与个体的特征或其他可能影响实验结果的因素完全独立。

解释变量“处理水平”与被遗漏的扰动项不相关，可避免遗漏变量偏差(omitted variable bias)或内生变量偏差(endogeneity bias)。

考虑以下回归模型：

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

x_i完全随机地决定。由于x_i与ε_i相互独立，故$\mathrm{Cov}(x_i, \varepsilon_i) = 0$，无论遗漏多少变量，OLS 都一致。

由于x_i与ε_i相互独立，故$\mathrm{E}(\varepsilon_i \mid x_1, \cdots, x_n) = 0$，也满足小样本理论的严格外生性假定，故 OLS 无偏。

在理想的随机实验中，X 对 y 的因果效应(causal effect)表现在条件期望的差别，即

$$\mathrm{E}(y \mid X = x) - \mathrm{E}(y \mid X = 0)$$

称为"处理效应"(treatment effect)。

例　x_i为服用药量，比如，$x_i=\{0,1,2\}$，而y_i为病情康复情况。

例　$x_i=\{0,1\}$表示是否参加过某一就业培训项目(job training program)，y_i为未来的就业状态。

如果$x_i=\{0,1\}$为虚拟变量，则方程的 OLS 估计量为

$$\hat{\beta}_{\text{OLS}}=\bar{y}_{\text{treat}}-\bar{y}_{\text{control}}$$

$\bar{y}_{\text{treat}}$为实验组的样本均值，$\bar{y}_{\text{control}}$为控制组的样本均值。

在回归方程中加入虚拟变量的效果就相当于给予实验组与控制组不同的截距项。而当$\{y_i\}$对常数项回归，系数估计值就是$\bar{y}$。因此，

$$\hat{\alpha}_{\text{OLS}} = \overline{y}_{\text{control}}, \quad \hat{\alpha}_{\text{OLS}} + \hat{\beta}_{\text{OLS}} = \overline{y}_{\text{treat}}$$

由于$\hat{\beta}_{\text{OLS}}$等于实验组均值与控制组均值之差，故称为“差分估计量”(differences estimator)。

例 班级规模是否影响学习成绩？美国田纳西州进行了为期四年的随机实验，将幼儿园至小学三年级的学生随机分为三组。第一组为普通班，每班 22-25 名学生；第二组为小班，每班 13-17 名学生；第三组也为小班，但配备一名教学助理。教师也随机分到这三类班级。

Table 2.2.1: Comparison of treatment and control characteristics in the Tennessee STAR experiment

	Students who entered STAR in kindergarten				
	Variable	Small	Regular	Regular/Aide	Joint P-value
1.	Free lunch	.47	.48	.50	.09
2.	White/Asian	.68	.67	.66	.26
3.	Age in 1985	5.44	5.43	5.42	.32
4.	Attrition rate	.49	.52	.53	.02
5.	Class size in kindergarten	15.10	22.40	22.80	.00
6.	Percentile score in kindergarten	54.70	48.90	50.00	.00

Notes: Adapted from Krueger (1999), Table 1. The table shows means of variables by treatment status. The P-value in the last column is for the F-test of equality of variable means across all three groups. All variables except attrition are for the first year a student is observed, The free lunch variable is the fraction receiving a free lunch. The percentile score is the average percentile score on three Stanford Achievement Tests. The attrition rate is the proportion lost to follow up before completing third grade.

Table 2.2.2: Experimental estimates of the effect of class-size assignment on test scores

Explanatory variable	(1)	(2)	(3)	(4)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian (1 = yes)	–	–	8.35	8.44
			(1.35)	(1.36)
Girl (1 = yes)	–	–	4.48	4.39
			(.63)	(.63)
Free lunch (1 = yes)	–	–	-13.15	-13.07
			(.77)	(.77)
White teacher	–	–	–	-.57
				(2.10)
Teacher experience	–	–	–	.26
				(.10)
Master's degree	–	–	–	-0.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

Note: Adapted from Krueger (1999), Table 5. The dependent variable is the Stanford Achievement Test percentile score. Robust standard errors that allow for correlated residuals within classes are shown in parentheses. The sample size is 5681.

18.3 引入更多的解释变量

在理想的随机实验条件下，OLS(即差分估计量)一致且无偏，但由于遗漏较多变量，ε_i的方差可能较大，OLS 可能效率不高。

引入某些遗漏变量，可改善此问题；也提供了检验x_i是否完全随机的机会。

假设引入的其他解释变量为$\{z_{i1}, \cdots, z_{iK}\}$，

$$y_i = \alpha + \beta x_i + \delta_1 z_{i1} + \cdots + \delta_K z_{iK} + \varepsilon_i$$

如果x_i完全随机，则$\{z_{i1}, \cdots, z_{iK}\}$应对$x_i$没有解释力。可把$x_i$对

$\{z_{i1}, \cdots, z_{iK}\}$回归，并检验此方程的整体显著性。

例 家中有电脑能否改善教育效果(educational outcome)？Fairlie and London (2012)在美国一个社区学院的 286 名新生中进行了随机实验。这些学生被随机分成两组，一组(实验组)得到免费电脑，而另一组为控制组；问卷跟踪调查两年，并从学院获取相关信息。

18.4 随机实验执行过程中可能出现的问题

在随机实验的执行过程中，可能出现的问题包括，“内部有效性问题”(internal validity)与“外部有效性问题”(external validity)。

1．内部有效性问题

(1) 未能完全随机分组(failure to randomize)：

比如，以姓氏字母在字母表的前半部或后半部来决定谁进入就业培训项目。但姓氏与种族有关，而种族又与就业机会有关。可把x_i对个体特征$\{z_{i1},\cdots,z_{iK}\}$回归，并检验该回归方程的整体显著性。

(2) 未能完全遵从实验设计(partial compliance)：

比如，被指定参加就业培训者却因故没来，而未被指定参加者却自行来参加。

可使用工具变量法来解决x_i的内生性问题，以“设计处理水平”(assigned treatment level，记为w_i)作为“实际处理水平”(actual treatment level，记为x_i)的工具变量。

(3) 中途退出实验(attrition)：

比如，参加就业培训的优秀者可能在项目进行过程中就找到工作而退出项目，造成选择性偏差。如果退出原因与实验无关(比如，去照顾家人)，则不会造成选择性偏差。

(4) 实验效应或霍桑效应(experimental effect 或 Hawthorne effect)：

参加实验本身可能改变个体的心理或行为，从而影响实验结果。对于药物疗效实验，可以通过使用双盲法来避免这种效应，称为“霍桑效应”。

(5) 样本过小：由于实验成本高，实验的样本容量可能较小。

2．外部有效性问题

(1) 样本的代表性不足(non-representative sample)：

让即将获释的囚犯参与就业培训，其结论可能难以推广到普通人群。又比如，项目参加者为自愿报名的义工，而义工的素质通常高于普通人。

(2) 小型实验的条件与大规模推广时的现实条件不同：

经济学家在肯尼亚进行实验，旨在推广蚊帐防止疟疾。一种观点认为，免费发放蚊帐能够最快地推广蚊帐。另一种观点认为，人们不会珍惜免费蚊帐，且免费发放使得蚊帐的长期供给变得困难。经济学家随机地对一些村庄提供免费蚊帐，对另外一些村庄有偿提供。

实验结果证明，免费提供蚊帐更为有效。但实验结果如果大规模推广仍然有效吗？反驳者指出，在那个实验地区，蚊帐的价值已广为人知；在实验过程中，蚊帐供给也有保证；这两点在大规模推广时未必成立。

(3) 一般均衡效应(general equilibrium effect)：

一个小型的就业培训项目可能不会改变社会上雇主的行为。大面积推广后，雇主可能减少由企业自行提供的员工培训，使得该项目的社会净福利减少。

(4) 自我选择效应：

在随机实验中，数据可能显示该项目并不有效。但在现实社会中，

预期收益最大的人们将最有积极性参加此项目。因此，这个项目可能在实际上有效。

18.5 自然实验

随机实验通常成本较高。而自然实验为自然发生(非为实验目的而发生)，几乎没有成本。

自然实验可分为两类。在第一类自然实验中，个体的分组或处理水平完全由自然实验所决定，可直接用 OLS 估计因果效应。

在第二类自然实验中，个体的分组或处理水平只是部分地由自然实验所决定，此时应以自然实验所带来的随机变动作为工具变量进行估计。

例（第一类自然实验） 最低工资对就业的影响。提高法定最低工资(minimum wage)在多大程度上会影响对低技能工人的需求？1992 年美国新泽西州通过法律提高最低工资，但相邻的宾夕法尼亚州最低工资却保持不变。两个州的雇主仿佛被随机地分配到实验组(新泽西州)与控制组(宾夕法尼亚州)。

例（第二类自然实验） 服兵役是否影响退役后的长期收入？Angrist (1990)考察了越战期间的参军者，当时美国对全国年轻男子以生日抽签方式进行征兵。抽签结果完全随机，但是否参军还取决于体检结果，而且有些人得到豁免，另一些人却自愿参军。

例（第二类自然实验） 经济增长对内战的影响。是否经济增长慢导致内战概率上升？但内战也会拖累经济增长，故存在内生性。Miguel et al (2004)使用外生的降雨量变化(rainfall variation)作为自然实验，考察 41 个非洲国家在 1981-1999 年间，经济增长对内战概率的因果作用。

非洲国家的经济比较依赖于自然降雨的农业，故经济增长与降雨量有关，但并不完全取决于降雨量。Miguel et al (2004)使用降雨量变化作为经济增长的 IV，发现经济低增长显著增加内战概率。

例 京杭大运河流经省份的人均 GDP 平均而言高于其他省份。这是否可以归功于京杭大运河对区域经济增长的促进作用？

18.6 双重差分法

实验效果常需要一段时间才能显现。考虑以下两期面板数据：

$$y_{it} = \alpha + \gamma D_t + \beta x_{it} + u_i + \varepsilon_{it} \quad (i = 1, \cdots, n;\ t = 1, 2)$$

D_t为实验期虚拟变量($D_t = 1$，如果 $t = 2$，实验后；$D_t = 0$，如果 $t = 1$，实验前)，u_i为不可观测的个体特征。

政策虚拟变量(policy dummy)

$$x_{it} = \begin{cases} 1, & 若 i \in 实验组, 且 t = 2 \\ 0, & 其他 \end{cases}$$

当$t = 1$时，实验组与控制组并未受到任何不同对待，x_{it}都等于 0。

当$t=2$时，实验组$x_{it}=1$，而控制组x_{it}依然为 0。

如果实验未能完全随机化(比如，观测数据)，则x_{it}可能与u_i相关，导致 OLS 不一致。

由于是面板数据，可对原方程进行一阶差分，以消掉u_i，

$$\Delta y_i = \gamma + \beta x_{i2} + \Delta \varepsilon_i$$

用 OLS 估计上式，可得一致估计。根据与差分估计量同样的推理：

$$\hat{\beta}_{\text{OLS}} = \Delta \overline{y}_{\text{treat}} - \Delta \overline{y}_{\text{control}} = (\overline{y}_{\text{treat, 2}} - \overline{y}_{\text{treat, 1}}) - (\overline{y}_{\text{control, 2}} - \overline{y}_{\text{control, 1}})$$

此法称为“双重差分估计量”(Difference-in-Differences estimator，

简记 DD)，记为$\hat{\beta}_{\text{DD}}$。双重差分估计量已剔除实验组与控制组“实验前差异”(pretreatment differences)的影响。

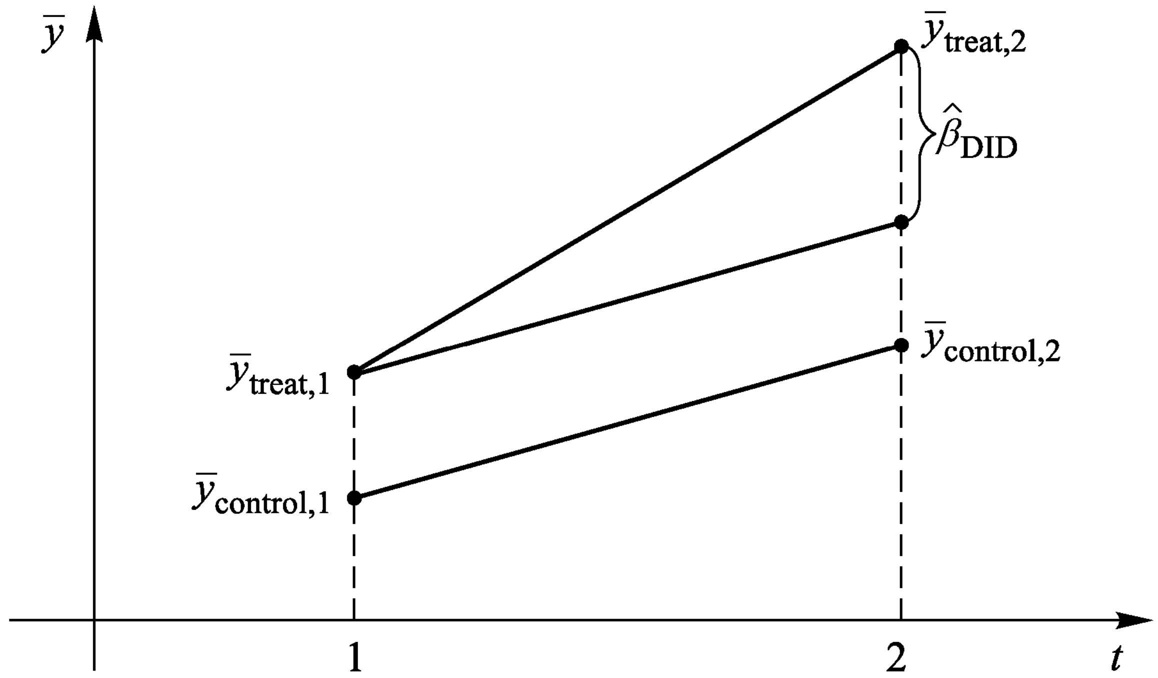

图 18.1　双重差分估计量示意图

对于双重差分估计量，也可引入其他解释变量$\{z_{i1},\cdots,z_{iK}\}$：

$$\Delta y_i = \gamma + \beta x_{i2} + \delta_1 z_{i1} + \cdots + \delta_K z_{iK} + \Delta\varepsilon_i$$

以Δy_i为被解释变量的双重差分法不适用于多期的数据。

回到以y_{it}为被解释变量的面板模型，暂时忽略其他解释变量，仍假设两期数据。则原方程与以下两期面板模型等价：

$$y_{it} = \beta_0 + \beta_1 G_i \cdot D_t + \beta_2 G_i + \gamma D_t + \varepsilon_{it} \quad (i=1,\cdots,n;\ t=1,2)$$

其中，G_i为实验组虚拟变量(G_i=1，个体 i 属于实验组；G_i=0，个体 i 属于控制组)；D_t为实验期虚拟变量(D_t=1，如果 t=2；D_t=0，如果 t=1)，互动项$G_i \cdot D_t = x_{it}$(取值为 1，若$i \in$实验组，且$t=2$；反

之，取值为 0)。

分组虚拟变量G_i刻画实验组与控制组本身的差异(即使不进行实验，也存在此差异)。

时间虚拟变量D_t刻画实验前后两期本身的差异(即使不进行实验，也存在此时间趋势)。

互动项$G_i \cdot D_t$才真正度量实验组的政策效应。

如有其他解释变量$\{z_{i1}, \cdots, z_{iK}\}$，可直接放入方程。

当 t =1 时，方程可以写为

$$y_{i1} = \beta_0 + \beta_2 G_i + \varepsilon_{i1}$$

当 t =2 时，方程可以写为

$$y_{i2} = \beta_0 + \beta_1 G_i \cdot D_2 + \beta_2 G_i + \gamma + \varepsilon_{i2}$$

两方程相减可得：

$$\Delta y_i = \gamma + \beta_1 G_i \cdot D_2 + (\varepsilon_{i2} - \varepsilon_{i1}) = \gamma + \beta_1 x_{i2} + \Delta\varepsilon_i$$

此方程与差分方程完全相同。对此方程进行 OLS 估计，得到的

$\hat{\beta}_1$(即互动项$G_i \times D_t$的系数)就是双重差分估计量。

面板形式的双重差分估计量，很容易推广到多期数据。

比如，共有 4 期数据，可估计如下方程：

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 G_i + \gamma_1 D2_t + \gamma_2 D3_t + \gamma_4 D4_t + \varepsilon_{it} \quad (i = 1, \cdots, n;\ t = 1, \cdots, 4)$$

其中，$D2_t, \cdots, D4_t$分别为对应于第 2-4 期的时间虚拟变量；政策虚拟变量x_{it}定义为

$$x_{it} = \begin{cases} 1, & \text{若}\, i \in \text{实验组}, \text{且}\, t \in \text{实验期} \\ 0, & \text{其他} \end{cases}$$

x_{it}系数的 OLS 估计值$\hat{\beta}_1$就是双重差分估计量。在两期数据中，x_{it}就是交叉项$G_i \times D_t$。

双重差分法的优点在于，同时控制了分组效应G_i(group-specific effects)与时间效应D_t(time-specific effects)。

18.7　三重差分法（选读）

18.8 观测数据的处理效应

在许多情况下，并没有随机实验或自然实验的数据，而只有观测数据(observational data)。

比如，政府提供的就业培训项目，完全由个人决定是否参与。

由于实际处理水平x_i存在自我选择，并非随机分组，故可能导致不一致的估计。

对于观测数据，须使用特别的方法来估计处理效应，详见第 28 章。