第四讲

虚拟变量回归模型

Econometrics

引子:男女大学生消费真有差异吗?

在对在校学生的消费行为进行的调查中,发现在校生的消费行为呈现多元化的结构。人际交往消费、手机类消费、衣着类消费、化妆品类消费、电脑类消费、旅游类消费占有较大的比例;而食品类消费、学习用品类消费不突显。

显然,男女生在消费上存在差异。为了了解男、女生的消费支出结构差异,应当如何建立模型?

面临的问题: 如何把男女生这样的非数量变量引入方程?

Econometrics

问题的一般性描述

在实际建模中，一些定性变量具有不可忽视的重要影响。例如，研究某个企业的销售水平，产业属性（制造业、零售业）、所有制（私营、非私营）、地理位置（东、中、西部）、管理者的素质、不同的收入水平等是值得考虑的重要影响因素，但这些因素共同的特征是定性描述的。

如何对非定量因素进行回归分析？

采用“虚拟变量”对定性变量进行量化一种思路。

Econometrics

第七讲 虚拟变量回归

本章主要讨论:

- 虚拟变量
- 虚拟解释变量的回归
- 虚拟被解释变量的回归(选讲, 不包括)

第一节 虚拟变量

本节基本内容:

- 基本概念
- 虚拟变量设置规则

Econometrics

一、基本概念

定量因素：可直接测度、数值性的因素。

定性因素：属性因素，表征某种属性存在与否的非数值性的因素。

基本思想：

直接在回归模型中加入定性因素存在诸多的困难（那些困难？），是否可将这些定性因素进行量化，以达到定性因素能与定量因素有着相同作用之目的。

虚拟变量的定义

计量经济学中，将取值为0和1的人工变量称为虚拟变量。虚拟变量也称：哑元变量、定性变量等等。通常用字母**D**或**DUM**加以表示（英文中虚拟或者哑元**Dummy**的缩写）。

对定性变量的量化可采用虚拟变量的方式实现。

Econometrics

二、虚拟变量设置规则

虚拟变量的设置规则涉及三个方面:

1. “0”和“1”选取原则

2. 属性（状态、水平）因素与设置虚拟变量数量的关系

3. 虚拟变量在回归分析中的角色以及作用等方面的问题

Econometrics

“0”和“1”选取原则

- 虚拟变量取“1”或“0”的原则，应从分析问题的目的出发予以界定。
- 从理论上讲，虚拟变量取“0”值通常代表比较的基础类型；而虚拟变量取“1”值通常代表被比较的类型。

“0”代表基期（比较的基础，参照物）；

“1”代表报告期（被比较的效应）。

例如，比较收入时考察性别的作用。当研究男性收入是否高于女性时，是将女性作为比较的基础（参照物），故有男性为“1”，女性为“0”。

例1

（1） $D=\begin{cases}1 & \text{男}\\ 0 & \text{女}\end{cases}$ （2） $D=\begin{cases}1 & \text{改革开放以后}\\ 0 & \text{改革开放以前}\end{cases}$

（3） $D_1=\begin{cases}1 & \text{天气阴}\\ 0 & \text{其　他}\end{cases}$ （4） $D_2=\begin{cases}1 & \text{天气雨}\\ 0 & \text{其　他}\end{cases}$

问题：

为何只选0、1，选2、3、4行吗？为什么？

Econometrics

属性的状态（水平）数与虚拟变量数量的关系

定性因素的属性既可能为两种状态，也可能为多种状态。例如，性别（男、女两种）、季节（4种状态），地理位置（东、中、西部），行业归属，所有制，收入的分组等。

如：$(D_1, D_2) = \begin{cases} (1,0) & \text{天气阴} \\ (0,1) & \text{天气雨} \\ (0,0) & \text{其　他} \end{cases}$

Econometrics

虚拟变量数量的设置规则

1. 若定性因素具有 m 个 $(m \geq 2)$相互排斥属性(或几个水平)，当回归模型有截距项时，只能引入 m-1个虚拟变量；

2. 当回归模型无截距项时，则可引入 m 个虚拟变量；否则，就会陷入“虚拟变量陷阱”。（为什么？）

Econometrics

一个例子(虚拟变量陷阱)

研究居民住房消费支出 Y_i 和居民可支配收入 X_i 之间的数量关系。回归模型的设定为：$Y_i = \alpha_0 + \beta_1 X_i + u_i$ （1）

现在要考虑城镇居民和农村居民之间的差异，如何办？

为了对 “城镇居民”、“农村居民”进行区分，分析各自在住房消费支出 Y_i 上的差异，设 $D_{1i}=1$ 为城镇；$D_{1i}=0$ 为农村，则模型为

$$Y_i = \alpha_0 + \beta_1 X_i + \alpha_1 D_1 + u_i \qquad (2)$$

(模型有截距，“居民属性”定性变量只有两个相互排斥的属性状态（$m=2$），故只设定一个虚拟变量。)

若对两个相互排斥的属性 “居民属性” ，仍然引入 $m=2$ 个虚拟变量，则有

$$D_{1i}=\begin{cases}1 & \text{城镇居民}\\ 0 & \text{农村居民}\end{cases} \qquad D_{2i}=\begin{cases}1 & \text{农村居民}\\ 0 & \text{城镇居民}\end{cases}$$

则模型（1）为

$$Y_i=\alpha_0+\beta_1 X_i+\alpha_1 D_1+\alpha_2 D_2+u_i \qquad (3)$$

则对任一家庭都有： $D_1+D_2=1$ $\quad D_1+D_2-1=0$ ，

即产生完全共线，陷入了“虚拟变量陷阱”。

“虚拟变量陷阱”的实质是：完全多重共线性。

虚拟变量在回归模型中的角色

虚拟变量既可作为被解释变量，也可作为解释变量，分别称其为虚拟被解释变量和虚拟解释变量。虚拟被解释变量的研究是当前计量经济学研究的前沿领域，如**MacFadden**、**Heckmen**等人的微观计量经济学研究，大量涉及到虚拟被解释变量的分析。本课程只是讨论虚拟解释变量的问题

Econometrics

第二节 虚拟解释变量的回归

本节基本内容:

- 加法类型
- 乘法类型
- 虚拟解释变量综合应用

Econometrics

在计量经济学中，通常引入虚拟变量的方式分为加法方式和乘法方式两种：即

$$Y_t = \alpha_0 + \beta X_t + u_t + \alpha_1 D$$

$$Y_t = \alpha + \beta_1 X_t + u_t + \beta_2 X_t D$$

原模型： $Y_i = \alpha + \beta X_i + u_i$

加法方式引入 $\alpha = \alpha_0 + \alpha_1 D$

乘法方式引入 $\beta = \beta_1 + \beta_2 D$

实质:加法方式引入虚拟变量改变的是截距；

乘法方式引入虚拟变量改变的是斜率。

Econometrics

一、加法类型

以加法方式引入虚拟变量时，主要考虑的问题是定性因素的属性和引入虚拟变量的个数。

分为四种情形讨论：

（1）解释变量只有一个定性变量而无定量变量，而且定性变量为两种相互排斥的属性；

（2）解释变量分别为一个定性变量（两种属性）和一个定量解释变量；

（3）解释变量分别为一个定性变量（两种以上属性）和一个定量解释变量；

（4）解释变量分别为两个定性变量（各自分别是两种属性）和一个定量解释变量；

思考：

四种加法方式引入虚拟变量会产生什么效应？

(1) 一个两种属性定性解释变量而无定量变量的情形

模型形式：$Y_i = f(D_i) + \mu_i \Rightarrow \alpha = \alpha_0 + \alpha_1 D_i$

例如：$Y_i = \alpha_0 + \alpha_1 D_i + \mu_i$

其中：$D_i = \begin{cases} 1 & \text{城市} \\ 0 & \text{农村} \end{cases}$ （比较的基础：农村）

那么：$E(Y_i \mid D_i = 1) = (\alpha_0 + \alpha_1)$

$E(Y_i \mid D_i = 0) = \alpha_0$

$Y_i = (\alpha_0 + \alpha_1) + \mu_i$ 城市

$Y_i = \alpha_0 + \mu_i$ 农村

(2) 一个定性解释变量（两种属性）和一个定量解释变量的情形

模型形式 $Y_i = f(D_i, X_i) + \mu_i \Rightarrow \alpha = \alpha_0 + \alpha_1 D_i$

例如：$Y_i = \alpha_0 + \alpha_1 D_i + \beta X_i + \mu_i$

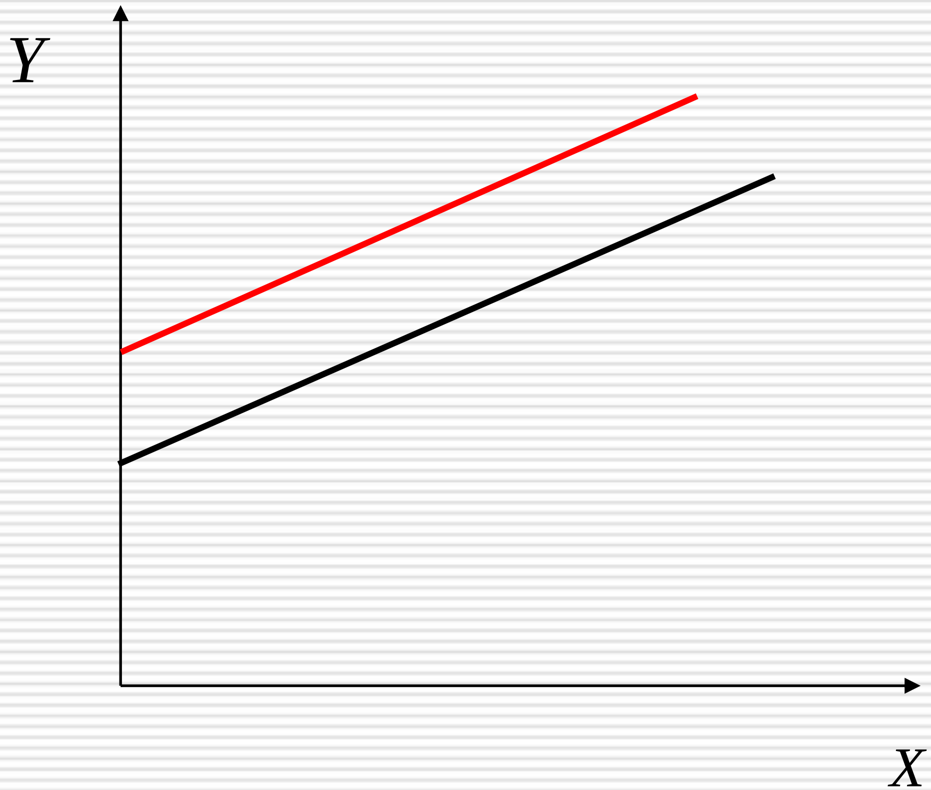
其中：Y—支出；X—收入；$D_i = \begin{cases} 1 & \text{城市} \\ 0 & \text{农村} \end{cases}$

$$E(Y_i \mid X_i, D_i = 1) = (\alpha_0 + \alpha_1) + \beta X_i$$

$$E(Y_i \mid X_i, D_i = 0) = (\alpha_0) + \beta X_i$$

$$Y_i = (\alpha_0 + \alpha_1) + \beta X_i + \mu_i \quad \text{城市}$$

$$Y_i = \alpha_0 + \beta X_i + \mu_i \quad \text{农村}$$



共同的特征：截距发生改变（？）

Econometrics

（3）一个定性解释变量（两种以上属性）和一个定量解释变量的情形

模型形式 $Y_i = f(X_i, D_1, D_2, \ldots) + \mu_i$

（如：民族有56种特性；季度有4种特性）

例如：啤酒售量Y、人均收入X、季度D；

$$Y_i = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta X_i + \mu_i$$

其中：$D_1 = \begin{cases} 1 & \text{一季度} \\ 0 & \text{其　它} \end{cases}$ $D_2 = \begin{cases} 1 & \text{二季度} \\ 0 & \text{其　它} \end{cases}$

$$D_3 = \begin{cases} 1 & \text{三季度} \\ 0 & \text{其　它} \end{cases}$$

Econometrics

一季度：$\mathrm{E}\left(Y_i \mid X_1, D_1=1, D_2=D_3=0\right)=(\alpha_0+\alpha_1)+\beta X_i$

二季度：$\mathrm{E}\left(Y_i \mid X_1, D_2=1, D_1=D_3=0\right)=(\alpha_0+\alpha_2)+\beta X_i$

三季度：$\mathrm{E}\left(Y_i \mid X_1, D_3=1, D_1=D_2=0\right)=(\alpha_0+\alpha_3)+\beta X_i$

四季度：$\mathrm{E}\left(Y_i \mid X_1, D_1=D_2=D_3=0\right)=\alpha_0+\beta X_i$

(基准：四季度)

(4) 两个定性解释变量（均为两种属性）和一个定量解释变量的情形

例：分析啤酒销售量Y受到人均收入X、季节D_1以及居民属性D_2的影响。

$$Y_i = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \beta X_i + \mu_i$$

其中：$D_1 = \begin{cases} 1 & \text{夏季} \\ 0 & \text{冬季} \end{cases}$　$D_2 = \begin{cases} 1 & \text{城市} \\ 0 & \text{农村} \end{cases}$

(比较的基础—冬季、农村)

Econometrics

夏季、城市居民

$$E\left(Y_i \mid X_i, D_1=1, D_2=1\right)=(\alpha_0+\alpha_1+\alpha_2)+\beta X_i$$

夏季、农村居民

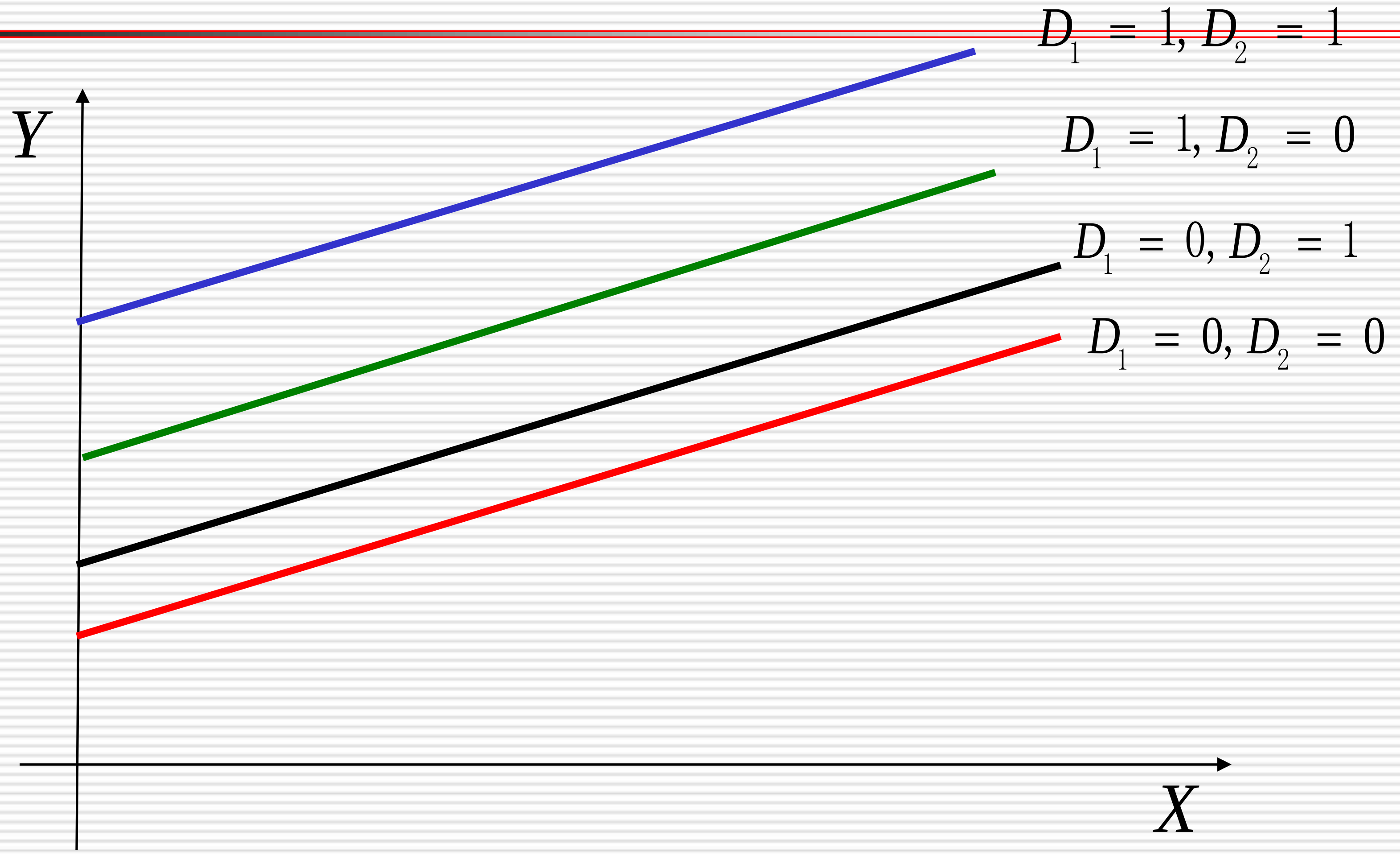
$$E\left(Y_i \mid X_i, D_1=1, D_2=0\right)=(\alpha_0+\alpha_1)+\beta X_i$$

冬季、城市居民

$$E\left(Y_i \mid X_i, D_1=0, D_2=1\right)=(\alpha_0+\alpha_2)+\beta X_i$$

冬季、农村居民

$$E\left(Y_i \mid X_i, D_1=0, D_2=0\right)=\alpha_0+\beta X_i$$



上述图形的前提条件是什么？

Econometrics

运用**OLS**得到回归结果，再用**t**检验讨论因素是否对模型有影响。

加法方式引入虚拟变量的一般表达式:

$$Y_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \ldots + \alpha_k D_{kt} + \beta X_t + u_t$$

基本分析方法：条件期望。

$$\mathrm{E}(Y_t / D_{1t}, D_{2t}, \ldots, D_{kt}) = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \ldots + \alpha_k D_{kt} + \beta X_t$$

加法方式引入虚拟变量的主要作用为：

1. 在有定量解释变量的情形下，主要改变方程截距；
2. 在没有定量解释变量的情形下，主要用于方差分析。

二、乘法类型

基本思想

以乘法方式引入虚拟变量时，是在所设立的模型中X_i，将虚拟解释变量与其它解释变量的乘积，作为新的解释变量出现在模型中，以达到其调整设定模型斜率系数的目的。或者将模型斜率系数表示为虚拟变量的函数，以达到相同的目的。

乘法引入方式:

（1）截距不变；

（2）截距和斜率均发生变化；

分析手段：仍然是条件期望。

(1) 截距不变的情形

模型形式：

$$Y_t = f\left(X_t, D_t X_t\right) + u_t \Rightarrow \alpha = \alpha, \beta = \beta_1 + \beta_2 D$$

例：研究消费支出 Y 受收入 X、年份状况 D 的影响

$$Y_t = \alpha + \beta_1 X_t + \beta_2 (D_t X_t) + \mu_t$$

其中：Y – 消费支出；X – 收入； $D_t = \begin{cases} 1 & \text{反常年份} \\ 0 & \text{正常年份} \end{cases}$

反常年份 $\mathrm{E}\left(Y_t \mid X_t, D_t = 1\right) = \alpha + (\beta_1 + \beta_2) X_t$

正常年份 $\mathrm{E}\left(Y_t \mid X_t, D_t = 0\right) = \alpha + \beta_1 X_t$

在正常年份的基础上进行比较，（只有斜率系数发生改变）。

Econometrics

（2）截距和斜率均发生变化

模型形式：

$$Y_i = f\left(X_t, D_t, D_t X_t\right) \Rightarrow \alpha = \alpha_0 + \alpha_1 D, \beta = \beta_1 + \beta_2 D$$

例，同样研究消费支出 Y 、收入 X 、年份状况 D 间的影响关系。

$$Y_t = \alpha_0 + \beta_1 X_t + \alpha_1 D_t + \beta_2 (D_t X_t) + \mu_t$$

其中： Y －消费支出； X －收入； $D_t = \begin{cases} 1 & \text{反常年份} \\ 0 & \text{正常年份} \end{cases}$

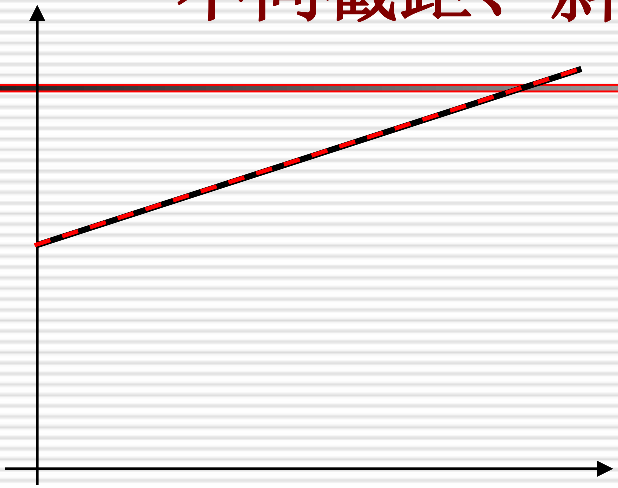
反常年份 $\mathrm{E}\left(Y_t \mid X_t, D_t = 1\right) = \left(\alpha_0 + \alpha_1\right) + (\beta_1 + \beta_2) X_t$

正常年份 $\mathrm{E}\left(Y_t \mid X_t, D_t = 0\right) = \alpha + \beta_1 X_t$

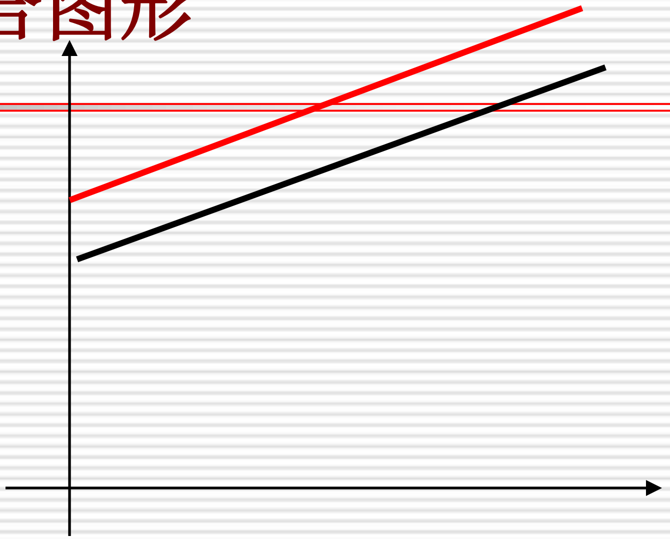
在正常年份基础上比较，截距和斜率系数都改变，为什么？

Econometrics

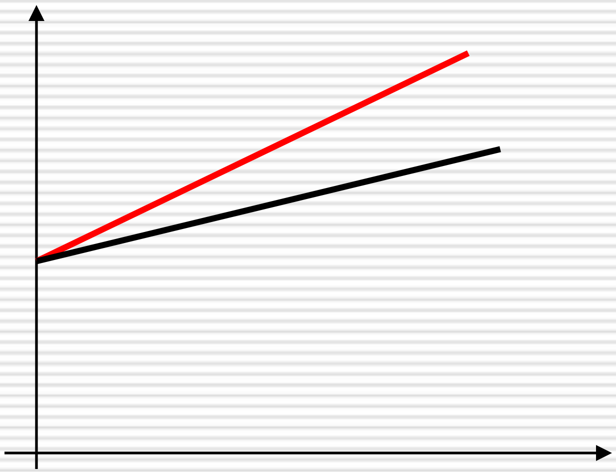
不同截距、斜率的组合图形



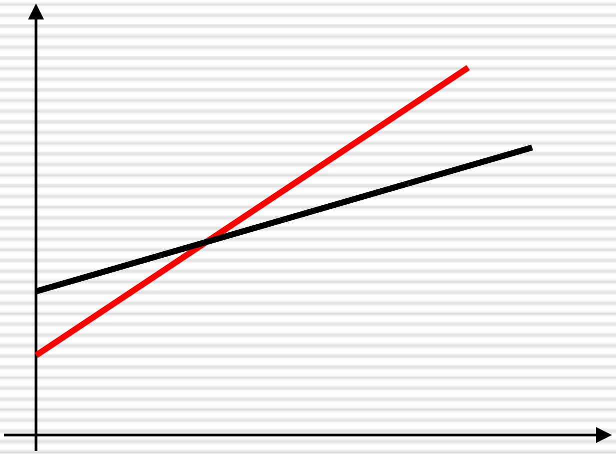
重合回归：截距斜率均相同



平行回归：截距不同斜率相同



共点回归：截距相同斜率不同



交叉（不同）回归：截距斜率均不同

Econometrics

三、虚拟解释变量综合应用

所谓综合应用是指将引入虚拟解释变量的加法方式、乘法方式进行综合使用。

基本分析方式仍然是条件期望分析。

本课主要讨论

（1）结构变化分析；

（2）交互效应分析；

（3）分段回归分析

Econometrics

(1) 结构变化分析

结构变化的实质是检验所设定的模型在样本期内是否为同一模型。显然，平行回归、共点回归、不同的回归三个模型均不是同一模型。

平行回归模型的假定是斜率保持不变（加法类型，包括方差分析）；

共点回归模型的假定是截距保持不变（乘法类型，又被称为协方差分析）；

不同的回归的模型的假定是截距、斜率均为变动的（加法、乘法类型的组合）。

Econometrics

例：比较改革开放前、后我国居民（平均）“储蓄—收入”总量关系是否发生了变化？

模型的设定形式为 ：

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \quad (1)$$

其中： Y_t为储蓄总额，X_t为收入总额。

$$D = \begin{cases} 1 & \text{改革开放后} \\ 0 & \text{改革开放前} \end{cases}$$

Econometrics

回归方程：

改革开放后 $\mathrm{E}(Y_t \mid X_t, D=1) = (\alpha_1+\alpha_2)+(\beta_1+\beta_2)X_t$ (2)

改革开放前 $\mathrm{E}(Y_t \mid X_t, D=0) = \alpha_1+\beta_1 X_t$ (3)

显然，只要α_2、β_2不同时为零，上述模型就能刻画改革开放前后我国居民储蓄收入模型结构是否发生变化。

问题：

1.本例中，平行、共点回归、不同的回归三模型的经济学背景解释是什么？

2.如何进行结构变化判断？

3.是否可对(2)、(3)分别进行 OLS 估计？为什么？

4.若分别对(2)、(3)进行 OLS 估计应注意什么？

(2）交互效应分析

交互作用:

一个解释变量的边际效应有时可能要依赖于另一个解释变量。为此，**Klein**和**Morgen**(1951)提出了有关收入和财产在决定消费模式上相互作用的假设。他们认为消费的边际倾向不仅依赖于收入，而且也依赖于财产的多少 ——较富有的人可能会有不同的消费倾向。

为了捕获该影响，设 $C = \alpha + \beta Y + u$。假设边际消费倾向 β 依赖于财产 Z。一个简单的表示方法就是 $\beta = \beta_1 + \beta_2 Z$。代入消费函数，有：

$$C = \alpha + \beta_1 Y + \beta_2 YZ + u$$

由于 YZ 捕获了收入和财产之间的相互作用而被称为交互作用项。

显然，刻画交互作用的方法，在变量为数量(定量)变量时， 是以乘法方式引入虚拟变量的。

Econometrics

例：是否发展油菜籽生产与是否发展养蜂生产的差异对农副产品总收益的影响研究。

模型设定为:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \qquad (1)$$

其中：Y_i（农副产品收益）；X_i（农副产品投入）

$$D_2 = \begin{cases} 1 & \text{发展油菜籽生产} \\ 0 & \text{其他} \end{cases}; \quad D_3 = \begin{cases} 1 & \text{发展养蜂生产} \\ 0 & \text{其他} \end{cases}$$

（1）式中，以加法形式引入虚拟变量暗含何假设?

（1）式以加法形式引入，暗含的假设为：

菜籽生产和养蜂生产是分别独立地影响农副品生产总收益。但是，在发展油菜籽生产时，同时也发展养蜂生产，所取得的农副产品生产总收益，可能会高于不发展养蜂生产的情况。即在是否发展油菜籽生产与养蜂生产的虚拟变量 D_{2i} 和 D_{3i} 间，很可能存在着一定的交互作用，且这种交互影响对被解释变量农副产品生产收益会有影响。

问题：如何刻画同时发展油菜籽生产和养蜂生产的交互作用？

基本思想:在模型中引入相关的两个变量的乘积。

区别之处在于，上页定义中的交互效应是针对数量变量，而现在是定性变量, 又应当如何处理?

Econometrics

为了反映交互效应，将（1）变为：

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{2i} D_{3i} + \beta X_i + u_i$$

同时发展油菜籽和养蜂生产：

$$Y_i = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i + u_i$$

发展油菜籽生产：

$$Y_i = (\alpha_1 + \alpha_2) + \beta X_i + u_i$$

发展养蜂生产：

$$Y_i = (\alpha_1 + \alpha_3) + \beta X_i + u_i$$

基础类型：

$$Y_i = \alpha_1 + \beta X_i + u_i$$

Econometrics

如何检验交互效应是否存在?

看$(D_{2i}D_{3i})$系数α_4对应的t值:

即检验: $\begin{cases} \mathrm{H_0}: \quad \alpha_4 = 0 \\ \mathrm{H_1}: \quad \alpha_4 \neq 0 \end{cases}$

若拒绝原假设,即交互效应对 Y 产生了影响(应该引入模型)。

(3) 分段回归分析

作用：提高模型的描述精度。

虚拟变量也可以用来代表数量因素的不同阶段。

分段线性回归就是类似情形中常见的一种。

一个例子： 研究不同时段我国居民的消费行为。

实际数据表明，1979年以前，我国居民的消费支出 Y_t 呈缓慢上升的趋势；从1979年开始，居民消费支出为快速上升趋势。

如何刻画我国居民在不同时段的消费行为？

基本思路：采用乘法方式引入虚拟变量的手段。显然，1979年是一个转折点，可考虑在这个转折点作为虚拟变量设定的依据。若设 X^* ＝1979，当 $t < X^*$ 时可引入虚拟变量。（为什么选择1979作为转折点？）

依据上述思路，有如下描述我国居民在不同时段消费行为模型：

$$Y_t = \beta_0 + \beta_1 t + \beta_2 (t - X^*) D + u_t$$

其中： $D = \begin{cases} 1 & t \geq X_t^* \\ 0 & t < X_t^* \end{cases}$ （t＝1955，1956，…，2004）

居民消费趋势方程：

1979年以前： $Y_t = \beta_0 + \beta_1 t + u_t$

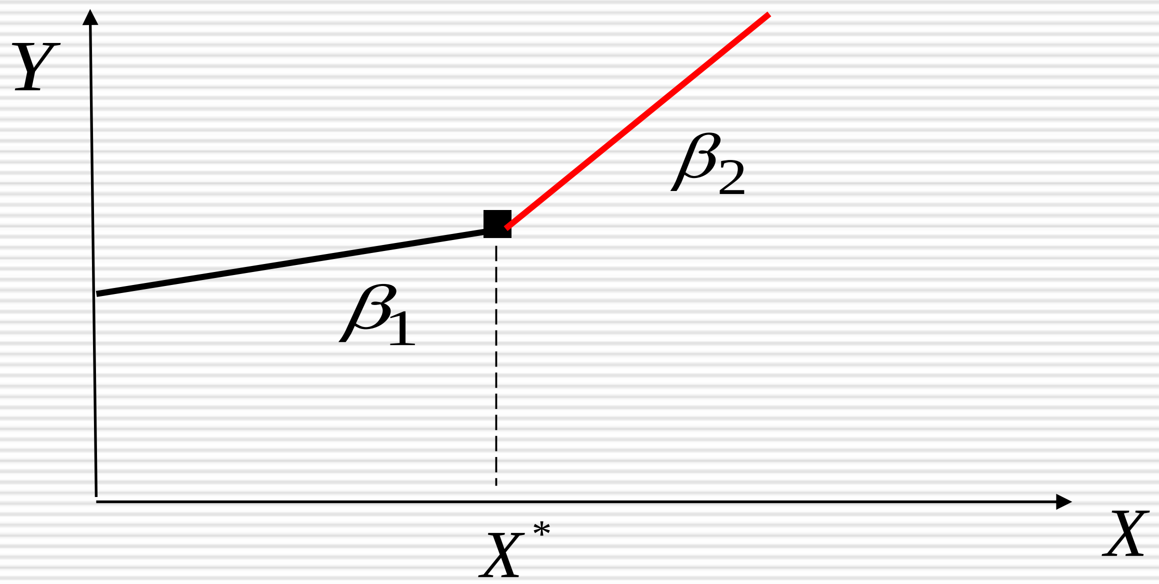
1979年以后： $Y_t = \beta_0 - \beta_2 X^* + (\beta_1 + \beta_2) t + u_t$

Econometrics

分析

1979年之前，回归模型的斜率为 β_1；

1979年之前，回归模型的斜率为 $\beta_1 + \beta_2$ ；



若统计检验表明，β_2显著不为零，则我国居民的消费行为在1979年前后发生了明显改变。

案例分析--虚拟变量的处理

- **一、实验基本原理**
- 对于定性数据或分类数据而言，通常并不能将其直接纳入模型中进行回归分析，因为这样的分析并不符合经济学理论，所以这时需要引入虚拟变量进行处理。一般情况下，如果分类变量总共有M类，为了避免多重共线性的出现，通常只引入M-1个虚拟变量。下面将会通过一个简单的例子，来介绍一下引入虚拟变量后，模型的实际变化。

例如，在如下的时间序列模型中：

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

假设在 t_1 时刻，经济结构发生了变动，这时就可以引入如下所示的虚拟变量 D 进行分析了。

$$D = \begin{cases} 1，\text{如果 } t \geq t_1 \\ 0，\text{如果 } t < t_1 \end{cases}$$

引入虚拟变量后的时间序列模型变为如下形式：

$$y_t = \beta_0 + \beta_1 x_t + \gamma D_t + \delta D_t x_t + \varepsilon_t$$

为了便于观察和比较，可以将此模型变化一下形式，即将虚拟变量取值带入，得到如下两个方程：

$$y_t = \begin{cases} \beta_0 + \beta_1 x_t + \varepsilon_t，\text{如果 } t < t_1 \\ (\beta_0 + \gamma) + (\beta_1 + \delta) x_t + \varepsilon_t，\text{如果 } t \geq t_1 \end{cases}$$

由此可以看出，当$\delta = 0$时，即只引入虚拟变量本身时，只会改变模型的截距；当$\gamma = 0$时，即只引入虚拟变量与其他解释变量的互动项时，只会改变模型的斜率；当$\delta \neq 0$ 且 $\gamma \neq 0$时，即两个部分都引入时，斜率和截距都会发生显著变化。虚拟变量就是通过这种方式使模型更加接近现实情况。

Econometrics

- 二、实验数据和实验内容
- 根据统计资料得到了中国1978—2006年的消费数据，变量主要包括：year=年份，c=人均消费（单位：元），y=人均国民收入（单位：元），c_ratio=消费收入比。完整的数据名为“consumption_china.dta”。

- 利用此数据，估计中国的消费函数，并引入虚拟变量，使得在1992年前后的模型截距和斜率都不相同。

三、实验操作指导

- 为了便于比较，首先生成整个时期中不含虚拟变量的消费函数方程，所使用到的命令为：
- regress c y
- 得到如图所示的回归结果，这个回归所形成的模型为c=188.588+0.3977y
- 如果认为在1992年，南巡讲话导致了经济结构的变动，这时需要引入虚拟变量将模型分成两段进行回归，步骤如下：
- 第一步，生成虚拟变量，所使用的命令为：
- generate dummy=0
- replace dummy=1 if year>=1992
- 在这组命令中，第一个命令的作用是生成虚拟变量dummy，使其值全部为0；第二个的命令的作用就是将1992年以后的dummy值替换为1，这时就完成了虚拟变量的设置。

Econometrics

- 第二步，生成虚拟变量dummy和解释变量y的互动项，所使用的命令为：
- generate dummy_y=dummy*y
- 这个命令的作用就是生成互动项dummy_y，使其值为变量dummy和变量y的乘积。
- 第三步，将虚拟变量纳入回归方程进行估计，所使用的命令为：
- regress c y dummy dummy_y
- 执行结果如图7.17所示，这时得到的模型为：

$$c=\begin{cases}12.9825+0.4996y，\text{如果 } year<1992\\(12.9825+553.4705)+(0.4996-0.1397)y，\text{如果 } year\geq 1992\end{cases}$$

- 这个模型是为了讲解虚拟变量的实际使用方法，暂不考虑某些系数不能通过检验的情况。通过引入虚拟变量发现，模型的截距和斜率都发生了变化。在用户实际研究过程中，可以根据需要引入虚拟变量，进行变斜率、变截距以及二者相结合的模型变化。

第四讲 小结

1. 虚拟变量是人工构造的取值为**0**和**1**的作为属性变量代表的变量。

2. 虚拟变量个数的设置有一定规则：在有截距项的模型中，若定性因素有 m个相互排斥的类型，只能引入 $m-1$个虚拟变量，否则会陷入所谓“虚拟变量陷阱”，产生完全的多重共线性。

3. 在计量经济模型中，加入虚拟解释变量的途径有两种基本类型：一是加法类型；二是乘法类型。以加法方式引入虚拟变量改变的是模型的截距；以乘法方式引入虚拟变量改变的是模型的斜率。

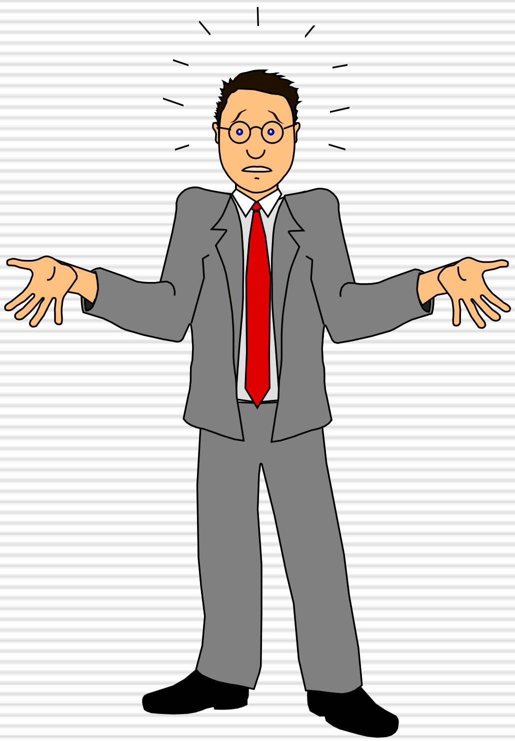
4. 解释变量只有一个分为两种相互排斥类型的定性变量而无定量变量的回归，称为方差分析模型。

5. 解释变量包含一个分为两种类型定性变量的回归时，只使用了一个虚拟变量；解释变量包含一个两种以上类型的定性变量的回归时，定性变量有 m 种类型，依据虚拟变量设置规则引入了 $m-1$ 个虚拟变量。

6. 解释变量包含两个（或 k 个）定性变量的回归中，可选用了两个（或 k 个）虚拟变量去表示，这并不会出现“虚拟变量陷阱”。

7. 以乘法形式引入虚拟解释变量的主要作用在于：对回归模型结构变化的检验；定性因素间交互作用的影响分析；分段线性回归等。

第四讲 结 束 了！



THANKS