

# 数据定价基础

2024-2025 学年短学期 数据要素市场

吴一航

yhwu\_is@zju.edu.cn

浙江大学计算机科学与技术学院

2024 年 7 月 3 日

# 消费者偏好

为了研究价格的决定，按照微观经济学的研究思路，我们需要从消费者偏好出发：如果你现在要组织一场活动，需要为活动购置一些钥匙扣和本子作为奖品，但你不确定买多少个钥匙扣和本子，这时你的心里会出现这样的声音：

- 比起 20 个钥匙扣和 5 个本子，我更喜欢 15 个钥匙扣和 10 个本子；
- 比起 15 个钥匙扣和 10 个本子，我更喜欢 10 个钥匙扣和 15 个本子；
- ...

# 消费者偏好

为了研究价格的决定，按照微观经济学的研究思路，我们需要从消费者偏好出发：如果你现在要组织一场活动，需要为活动购置一些钥匙扣和本子作为奖品，但你不确定买多少个钥匙扣和本子，这时你的心里会出现这样的声音：

- 比起 20 个钥匙扣和 5 个本子，我更喜欢 15 个钥匙扣和 10 个本子；
- 比起 15 个钥匙扣和 10 个本子，我更喜欢 10 个钥匙扣和 15 个本子；
- ...

在微观经济学中，我们假定消费者在面临选择时，通常首先会对可选方案进行排序，然后从中选择最喜欢的方案。这种排序就体现了消费者的**偏好 (preference)**。

# 消费者偏好 (Cont'd)

为了简化讨论，我们首先研究市场中只有两个商品的情形．我们把消费者的选择目标称为消费束，记为  $\mathbf{x} = (x_1, x_2)$ ，其中  $x_1$  和  $x_2$  分别表示消费者购买的两种商品的数量．任意给定两个消费束  $\mathbf{x} = (x_1, x_2)$  和  $\mathbf{y} = (y_1, y_2)$ ，我们可以用偏好关系 (**preference relation**)  $\mathbf{x} \succeq \mathbf{y}$  表示“ $\mathbf{x}$  至少和  $\mathbf{y}$  一样好”．我们可以从  $\succeq$  中推导出另外两个偏好关系：

- ① 严格偏好关系 (**strict preference**)  $\succ$ ，定义为

$$\mathbf{x} \succ \mathbf{y} \Leftrightarrow \mathbf{x} \succeq \mathbf{y} \text{ 且 } \mathbf{y} \not\succeq \mathbf{x}$$

我们将其读作“ $\mathbf{x}$  比  $\mathbf{y}$  好”．

- ② 无差异关系 (**indifference**)  $\sim$ ，定义为

$$\mathbf{x} \sim \mathbf{y} \Leftrightarrow \mathbf{x} \succeq \mathbf{y} \text{ 且 } \mathbf{y} \succeq \mathbf{x}$$

我们将其读作“ $\mathbf{x}$  和  $\mathbf{y}$  一样好”．

# 消费者偏好 (Cont'd)

在大部分经济理论中，经济学家都假设个人偏好是理性的 (rational). 这个假设体现在关于偏好关系  $\succeq$  的两个基本假设上：

- ① 完备性 (completeness): 对于任意两个消费束  $x$  和  $y$ , 要么  $x \succeq y$ , 要么  $y \succeq x$ .
- ② 传递性 (transitivity): 对于任意三个消费束  $x$ 、 $y$  和  $z$ , 如果  $x \succeq y$  且  $y \succeq z$ , 那么  $x \succeq z$ .

# 消费者偏好 (Cont'd)

在大部分经济理论中，经济学家都假设个人偏好是理性的 (rational). 这个假设体现在关于偏好关系  $\succeq$  的两个基本假设上：

- ① **完备性 (completeness)**: 对于任意两个消费束  $x$  和  $y$ , 要么  $x \succeq y$ , 要么  $y \succeq x$ .
- ② **传递性 (transitivity)**: 对于任意三个消费束  $x$ 、 $y$  和  $z$ , 如果  $x \succeq y$  且  $y \succeq z$ , 那么  $x \succeq z$ .

完备的这个假设，是说个人在任意两个可能的备选物上都有明确的偏好. 传递性直接触及理性这个概念的本质. 决策者在决策时将备选物两两比较，形成一个比较链条，传递性意味着决策者在这个比较链条上的偏好不可能是循环的. 例如，不可能出现下列这样的偏好：认为苹果至少和香蕉一样好，香蕉至少和橘子一样好，但橘子又比苹果好.

# 消费者偏好 (Cont'd)

根据偏好关系  $\succeq$  的理性假设, 我们不难推导出以下结论:

- ① 反身性 (reflexivity): 对于任意消费束  $x$ ,  $x \succeq x$ ,  $x \sim x$ , 但  $x \not\succ x$ .
- ② 传递性 (transitivity)
  - ① 如果  $x \succ y$  且  $y \succeq z$ , 那么  $x \succ z$ ;
  - ② 如果  $x \sim y$  且  $y \sim z$ , 那么  $x \sim z$ .

结论的证明留作练习.

# 孔多塞悖论 (Condorcet paradox)

然而，理性偏好很多时候并非合理的假定，例如一些明显不满足传递性的行为是很有趣的。考虑某个三口之家通过少数服从多数的投票机制进行决策。令  $M$ 、 $D$  和  $C$  分别表示妈妈、爸爸和孩子。他们面临的备选物是星期五晚上是看歌剧 ( $O$ )、摇滚演唱会 ( $R$ ) 还是花样滑冰表演 ( $I$ )。这三个人都是理性的：

- $O \succ_M R \succ_M I$ ;
- $I \succ_D O \succ_D R$ ;
- $R \succ_C I \succ_C O$ .

其中  $\succ_M$ 、 $\succ_D$  和  $\succ_C$  分别为这三个人的理性严格偏好关系。假设投票表决的分别是：歌剧 ( $O$ ) 对演唱会 ( $R$ )，演唱会 ( $R$ ) 对滑冰表演 ( $I$ ) 以及滑冰表演 ( $I$ ) 对歌剧 ( $O$ )。则投票结果为： $O$  战胜  $R$ ； $R$  战胜  $I$ ； $I$  战胜  $O$ 。因此，这个家庭的偏好为  $O \succ R \succ I \succ O$ ，显然不满足传递性。

事实上这个例子说明的非传递性，称为康多塞悖论 (Condorcet paradox)，是集体决策理论中的一个重要难题。



# 从偏好到效用函数

事实上，有了偏好之后，我们就可以衡量物品的价格了

- 事实上，我们只需要询问一个诚实的人，例如“你觉得 1 斤苹果和多少钱带给你的偏好是无差异的”，通过这样的询问，我们就可以得到一个消费者可以接受的价格
- 然而，这样的做法不够简洁美观，如果我们为消费者对 1 斤苹果的“满意程度”一个常数，那么我们就可以通过比较一些代表满意程度的常数就能得到消费者的偏好
- 效用函数就是将消费束映射到满意程度的函数：设  $\mathbf{x}$  是一个消费束，则  $u(\mathbf{x})$  是一个实数，表示消费者对  $\mathbf{x}$  的满意程度
  - 例如  $u(1 \text{ 斤苹果}, -6 \text{ 元钱}) = 0$ ，表示 1 斤苹果和 6 元钱对消费者来说是无差异的

# 从偏好到效用函数 (Cont'd)

问题：是否存在能完美表示偏好的效用函数呢？即满足

$$u(\mathbf{x}) \geq u(\mathbf{y}) \Leftrightarrow \mathbf{x} \succeq \mathbf{y}$$

## 定理：连续效用函数存在性

若理性偏好关系  $\succeq$  满足连续性，则存在一个连续的效用函数  $u(\mathbf{x})$ ，使得对于任意两个消费束  $\mathbf{x}$  和  $\mathbf{y}$ ，有

$$u(\mathbf{x}) \geq u(\mathbf{y}) \Leftrightarrow \mathbf{x} \succeq \mathbf{y}$$

什么是偏好的连续性呢？

## 定义：偏好的连续性

偏好关系  $\succeq$  是连续的，如果对于任意两列消费束序列  $\{\mathbf{x}_n\}$  和  $\{\mathbf{y}_n\}$ ，如果  $\mathbf{x}_n \succeq \mathbf{y}_n$  对所有  $n$  成立，且  $\mathbf{x}_n \rightarrow \mathbf{x}$  且  $\mathbf{y}_n \rightarrow \mathbf{y}$ ，则有  $\mathbf{x} \succeq \mathbf{y}$ 。

# 从偏好到效用函数 (Cont'd)

效用函数的注意点：序数效用与基数效用

- 效用函数做单调变换仍然可以代表同一偏好关系

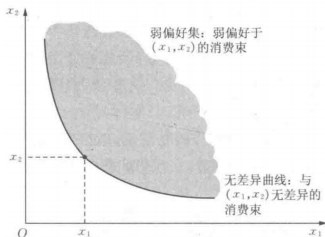
# 从偏好到效用函数 (Cont'd)

效用函数的注意点：序数效用与基数效用

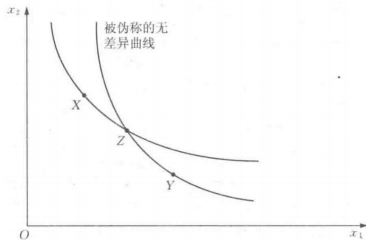
- 效用函数做单调变换仍然可以代表同一偏好关系

如何构建效用函数：无差异曲线

- 边际替代率（斜率绝对值）递减规律
- 两条无差异曲线不能相交



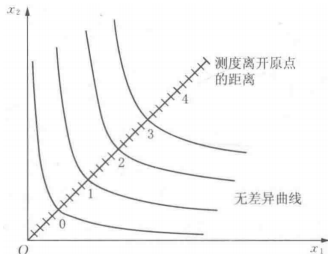
绘成阴影的区域由所有这样的消费束组成，这些消费束至少与消费束  $(x_1, x_2)$  一样好。



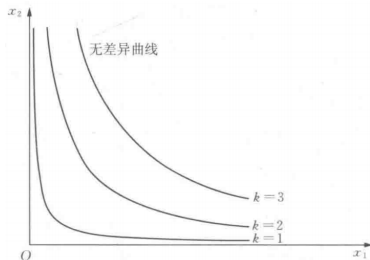
如果相交了，X、Y和Z相互之间都会无差异，因而也就不可能位于不同的无差异曲线上。

# 从偏好到效用函数 (Cont'd)

如何构建效用函数：无差异曲线



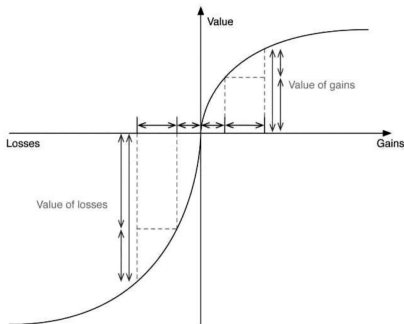
画一条对顶线, 用沿这条线测度的每条无差异曲线离原点的距离给每条无差异曲线标记数字。



对应于不同  $k$  值的无差异曲线  $k = x_1 x_2$ 。

# 常用效用函数

- 柯布-道格拉斯效用函数:  $u(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$
- 拟线性效用函数:  $u(\mathbf{x}, p) = v(\mathbf{x}) - p$ 
  - 表示通过价格  $p$  购买了消费束  $\mathbf{x}$  后的效用
  - 假定每一元钱的效用是单位 1
  - 日后最常见的效用函数
- 冯诺伊曼-摩根斯坦效用函数:
 
$$u(px + (1-p)y) = pu(\mathbf{x}) + (1-p)u(\mathbf{y})$$
  - 如果有一个彩票，购买后有  $p$  的概率获得  $\mathbf{x}$ ，有  $1-p$  的概率获得  $\mathbf{y}$ ，那么购买这个彩票的效用等于购买  $\mathbf{x}$  和购买  $\mathbf{y}$  的效用的加权平均
  - 风险厌恶与损失厌恶
    - 风险厌恶:
 
$$u(px + (1-p)y) < pu(\mathbf{x}) + (1-p)u(\mathbf{y})$$
    - 损失厌恶: 损失给投资者带来的痛苦远远大于等量收益给投资带来的安慰，所以投资者为避免损失必须要求风险资产有较高的期望收益



# 边际效用递减规律

炎炎夏日，如果你想买一些冰激凌解暑，你可能会思考吃几个冰激凌对你而言最满足，于是你计算了一下大致的效用值如下表：

数量	1	2	3	4	5	6
效用	5	12	16	18	15	10

# 边际效用递减规律

炎炎夏日，如果你想买一些冰激凌解暑，你可能会思考吃几个冰激凌对你而言最满足，于是你计算了一下大致的效用值如下表：

数量	1	2	3	4	5	6
效用	5	12	16	18	15	10

## 边际效用

边际效用是指消费者对某种物品的消费量每增加一单位所增加的额外满足程度。

数量	1	2	3	4	5	6
边际效用	5	7	4	2	-3	-5



# 边际效用递减规律

炎炎夏日，如果你想买一些冰激凌解暑，你可能会思考吃几个冰激凌对你而言最满足，于是你计算了一下大致的效用值如下表：

数量	1	2	3	4	5	6
效用	5	12	16	18	15	10

## 边际效用

边际效用是指消费者对某种物品的消费量每增加一单位所增加的额外满足程度。

数量	1	2	3	4	5	6
边际效用	5	7	4	2	-3	-5

在一定时间内，随着消费某种商品数量的不断增加，消费者从中得到的总效用是在增加的，但是以递减的速度增加的，即边际效用是递减的；当商品消费量达到一定程度后，总效用达到最大值，如果继续增加消费，总效用不但不会增加，反而会逐渐减少，此时边际效用变为负数。

# 消费者需求曲线

在预备工作结束后，我们可以开始研究如何推导消费者的需求曲线，即我们要研究需求与价格之间的关系。

# 消费者需求曲线

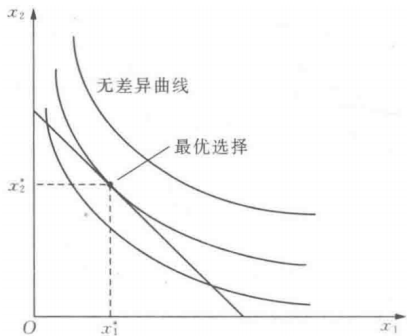
在预备工作结束后，我们可以开始研究如何推导消费者的需求曲线，即我们要研究需求与价格之间的关系。

为简化讨论，我们仍考虑只有两个商品的情形。设消费者的效用函数为  $u(x_1, x_2)$ ，价格分别为  $p_1$  和  $p_2$ ，消费者的收入为  $p$ 。这里我们强调了消费者的收入，这是因为在消费时消费者都有**预算约束 (budget constraint)**，因此我们的目标可以写为

$$\begin{aligned} \max_{x_1, x_2} \quad & u(x_1, x_2) \\ \text{s.t.} \quad & p_1 x_1 + p_2 x_2 \leq p \end{aligned}$$

这里我们使用了经济学中的基本原则，即**效用最大化原则 (principle of utility maximization)**：理性人都会选择使自己的效用最大化的行动选择，因为实际上效用最大化的选择就是自己最偏好的选择。

# 消费者需求曲线 (Cont'd)



最优消费位置处在无差异曲线与预算线的切点上。

- 最上方的无差异曲线对应额消费束在当前预算约束下无法购买
- 最下方的无差异曲线对应额消费束在当前预算约束下并非最优选择
- 最优选择是在无差异曲线与预算约束线的切点处（因此预算必须用完，约束取等号）
- 根据这一特点我们可以解出消费者的需求函数，即消费者需求与价格之间的关系

# 消费者需求曲线 (Cont'd)

## 例

设消费者需求函数为  $u(x_1, x_2) = x_1^\alpha x_2^\beta$ ，两种商品价格分别为  $p_1$  和  $p_2$ ，消费者的收入为  $p$ 。求消费者的需求函数。

- 结果为  $x_1 = \frac{\alpha}{\alpha + \beta} \frac{p}{p_1}$ ,  $x_2 = \frac{\beta}{\alpha + \beta} \frac{p}{p_2}$
- 可以利用预算约束化为一元函数极值问题求解，也可以使用拉格朗日乘子法求解
- 注意这里的效用函数没有考虑金钱效用

# 消费者需求曲线 (Cont'd)

## 例

设消费者需求函数为  $u(x_1, x_2) = x_1^\alpha x_2^\beta$ ，两种商品价格分别为  $p_1$  和  $p_2$ ，消费者的收入为  $p$ 。求消费者的需求函数。

- 结果为  $x_1 = \frac{\alpha}{\alpha + \beta} \frac{p}{p_1}$ ,  $x_2 = \frac{\beta}{\alpha + \beta} \frac{p}{p_2}$
- 可以利用预算约束化为一元函数极值问题求解，也可以使用拉格朗日乘子法求解
- 注意这里的效用函数没有考虑金钱效用

将这样的基本思想推广，我们可以得到一般情况下的消费者需求函数。

## 需求定律

在其他条件不变的情况下，商品的需求量与价格之间成反方向变动的关系，即价格上涨，需求量减少；价格下降，需求量增加。

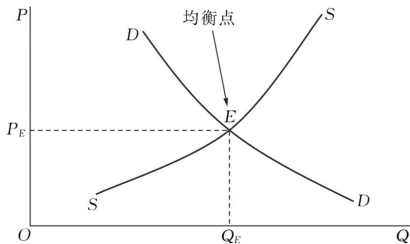
当然也存在一些例外，例如吉芬商品，但这里不做讨论。

# 供给曲线与市场出清

市场中除了消费者之外，还有一方是生产商品的厂商。厂商的决策基础也是自身的效用最大化，更具体而言则是利润最大化。我们略去对厂商的讨论，直接给出供给定律：

## 供给定律

对于正常商品来说，在其他条件不变的情况下，商品价格与需求量之间存在着正方向的变动关系，即一种商品的价格上升时，这种商品的供给量就会增加，相反，价格下降时供给量减少，这就是供给定律。



- 竞争市场与竞争性均衡
- 市场出清 (market clearing) : 市场机制能够自动地消除超额供给或超额需求，市场在短期内自发地趋于供给等于需求的均衡状态

# 福利经济学第一定理

市场中每个人都选择自私地最大化自身利益，能实现有效率的资源配置吗？

## 福利经济学第一定理

在竞争市场中，当市场供求达到均衡时，市场资源配置是帕累托有效率（Pareto efficiency）的。

- ① 帕累托有效率的含义之后解释，现在可以认为是一种有效率的资源配置方式
- ② 福利经济学第一定理的条件很强，要求完全竞争、完全信息、无交易成本、无外部性、无规模经济
- ③ 亚当·斯密“看不见的手”的数学解释



# 帕累托最优 (Pareto optimality)

## 帕累托最优

如果一个经济体的资源配置不能通过改变分配而使任何人的福利提高而不损害其他人的福利，那么这种资源配置就是**帕累托最优 (Pareto optimality)** 或**帕累托有效率 (Pareto efficiency)** 的。

- ① 帕累托最优的本质是一个多目标优化问题，目标最大化每个人的效用，最终优化到一个只能通过损害其他人的效用来提高某个人的效用的状态
- ② 帕累托最优不是社会福利最大化 (social welfare maximization)
  - 社会福利最大化可以简单看作每个人的效用之和最大化
  - 帕累托最优可能存在如下情况：损害一个人的效用可以让他人效用提高很多，这样帕累托最优的分配显然不是社会福利最大化的分配
  - 帕累托最优的点可以是无穷的（接下来分析），但社会福利最大化的一般不是

# 帕累托最优 (Pareto optimality) (Cont'd)

我们回到第一节牛肉和土豆的例子，这一经济的参与人简记为 A 和 B，假设 A 和 B 开始拥有的土豆（记为商品 1）和牛肉（记为商品 2）的数量分别为  $w_A^1, w_A^2$  和  $w_B^1, w_B^2$ 。

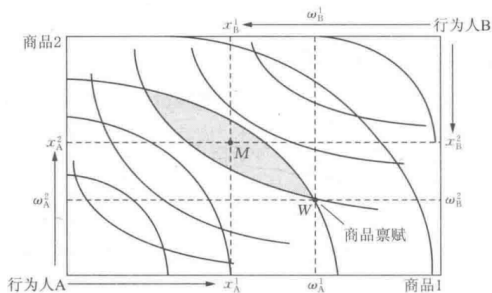
- 经济学中称初始拥有的商品为禀赋 (endowment)
- A 和 B 可以交换商品，交换后 A 和 B 拥有的商品 1 和 2 的数量分别为  $x_A^1, x_A^2$  和  $x_B^1, x_B^2$ 
  - 注意数量应当满足

$$x_A^1 + x_B^1 = w_A^1 + w_B^1$$

$$x_A^2 + x_B^2 = w_A^2 + w_B^2$$

- 回到第一章遗留的问题，二者交换牛肉和土豆到什么配置下会停止呢？

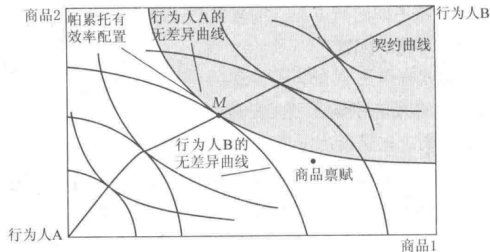
# 帕累托最优 (Pareto optimality) (Cont'd)



方框的长用以度量经济中商品1的总量,方框的宽用以度量商品2的总量。行为人A的消费选择以左下角为原点度量,而行为人B的消费选择则以右上角为原点度量。

- 我们从禀赋  $W$  出发, A 的境况可以改善的区域是  $W$  所在无差异曲线的右上方, B 的境况可以改善的区域是  $W$  所在无差异曲线的左下方
- 两个区域相交区域就是可以改进的区域,例如可以向  $M$  点移动(产生帕累托改进)
- 在  $M$  点可以做类似的分析,可以找到  $M$  点的改进区域,直到没有帕累托改进为止

# 帕累托最优 (Pareto optimality) (Cont'd)



- 显然，只有当改进到某个点时，A 和 B 的无差异曲线相切，才无法再找到帕累托改进，如图中的  $M$  点
- 我们可以发现，这样的帕累托最优点一定是有无穷多的，将它们连起来得到的曲线，我们称为**契约曲线 (contract curve)**
- 由此我们可以解答第一节的问题：A 和 B 的交换会在契约曲线上的某个点停止，这个点取决于双方的初始禀赋以及双方的谈判能力

# 传统理论的局限与数据定价的挑战

## ① 从市场角度看

- a. 到目前为止的传统理论研究的都是竞争市场，即价格都是市场决定的，每个参与人只是价格的接受者
- b. 我们不排除同质数据的存在，但很多数据都是企业等团体自己产生的，因此它们使得市场不是完全竞争的，因此如何实现数据定价是一个重要的问题
- c. 因此传统定价参考市场定价的方法失效

# 传统理论的局限与数据定价的挑战

## ① 从市场角度看

- a. 到目前为止的传统理论研究的都是竞争市场，即价格都是市场决定的，每个参与人只是价格的接受者
- b. 我们不排除同质数据的存在，但很多数据都是企业等团体自己产生的，因此它们使得市场不是完全竞争的，因此如何实现数据定价是一个重要的问题
- c. 因此传统定价参考市场定价的方法失效

## ② 从需求角度看

- a. 同一份数据的用途可能有很多，并且同一份数据在不同用途下的表现也不一定相同，因此买家的效用很难确定
- b. 因此从买家的偏好是很难确定的，这也导致市场是不完全信息的，与传统理论的假设不符

# 传统理论的局限与数据定价的挑战

## ① 从市场角度看

- a. 到目前为止的传统理论研究的都是竞争市场，即价格都是市场决定的，每个参与人只是价格的接受者
- b. 我们不排除同质数据的存在，但很多数据都是企业等团体自己产生的，因此它们使得市场不是完全竞争的，因此如何实现数据定价是一个重要的问题
- c. 因此传统定价参考市场定价的方法失效

## ② 从需求角度看

- a. 同一份数据的用途可能有很多，并且同一份数据在不同用途下的表现也不一定相同，因此买家的效用很难确定
- b. 因此从买家的偏好是很难确定的，这也导致市场是不完全信息的，与传统理论的假设不符

解决方案：以上问题都可以很自然地在博弈论的框架下解决，这也是我们下面几次课的核心。

# 传统理论的局限与数据定价的挑战

*They have a high and irrecoverable fixed cost (producing the data is expensive) and a very low variable cost (copying the data is almost free)*

## ③ 从供给角度看

- a. 数据具有零成本复制的特点，额外出售一单位的数据的边际成本几乎为零（不考虑隐私成本），根据传统理论，似乎数据应该无限复制出售，但真的无限出售后数据价格一定会降低，传统理论失效



# 传统理论的局限与数据定价的挑战

*They have a high and irrecoverable fixed cost (producing the data is expensive) and a very low variable cost (copying the data is almost free)*

## ③ 从供给角度看

- a. 数据具有零成本复制的特点，额外出售一单位的数据的边际成本几乎为零（不考虑隐私成本），根据传统理论，似乎数据应该无限复制出售，但真的无限出售后数据价格一定会降低，传统理论失效
- b. 零成本复制使得数据**外部性 (externalities)** 比一般商品强
  - 外部性指一个人或一群人的行动和决策使另一个人或一群人受损或受益的情况。经济外部性是经济主体（包括厂商或个人）的经济活动对他人和社会造成的非市场化的影响。即社会成员（包括组织和个人）从事经济活动时其成本与后果不完全由该行为人承担
  - 例如你的竞争厂家购买了某份利于提高产量的重要数据，但你没有买，尽管你没有参与市场，但你的利润可能会受到影响
  - 零成本复制使得数据出售更容易，人们更容易受到外部性的影响

# 传统理论的局限与数据定价的挑战

*They have a high and irrecoverable fixed cost (producing the data is expensive) and a very low variable cost (copying the data is almost free)*

## ③ 从供给角度看

- a. 数据具有零成本复制的特点，额外出售一单位的数据的边际成本几乎为零（不考虑隐私成本），根据传统理论，似乎数据应该无限复制出售，但真的无限出售后数据价格一定会降低，传统理论失效
- b. 零成本复制使得数据**外部性 (externalities)** 比一般商品强
  - 外部性指一个人或一群人的行动和决策使另一个人或一群人受损或受益的情况。经济外部性是经济主体（包括厂商或个人）的经济活动对他人和社会造成的非市场化的影响。即社会成员（包括组织和个人）从事经济活动时其成本与后果不完全由该行为人承担
  - 例如你的竞争厂家购买了某份利于提高产量的重要数据，但你没有买，尽管你没有参与市场，但你的利润可能会受到影响
  - 零成本复制使得数据出售更容易，人们更容易受到外部性的影响

解决方案：从面向成本定价转向面向需求定价，接下来会介绍相关的定价方法。