目录

[路径/后门路径 2](#_Toc128908269)

[有向无环图 2](#_Toc128908270)

[变量的类型与因果结构 2](#_Toc128908271)

[路径 5](#_Toc128908272)

[后门路径 vs 前门路径 5](#_Toc128908273)

[开放路径 vs 闭塞路径 5](#_Toc128908274)

[直接路径 vs 间接路径 6](#_Toc128908275)

[偏倚路径 vs 混杂路径 6](#_Toc128908276)

[混杂与偏倚控制 7](#_Toc128908277)

# 路径/后门路径

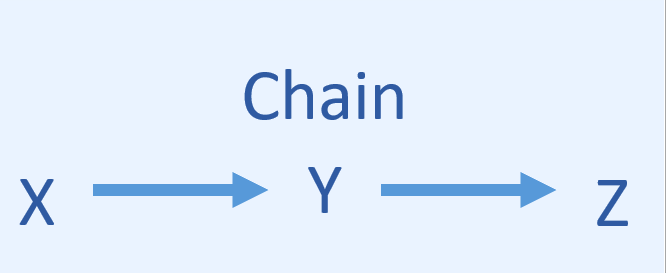
## 有向无环图

有向无环图(directed acyclic graph,DAG),又叫因果图,是结构因果模型(Structural causal model,SCM)理论的基石。以图形化的方法表示研究假设,指导因果推断，已逐渐被应用于流行病学研究。关于因果图的详细介绍可参考相关文献(见文末)。本文主要介绍几个要点。

## 变量的类型与因果结构

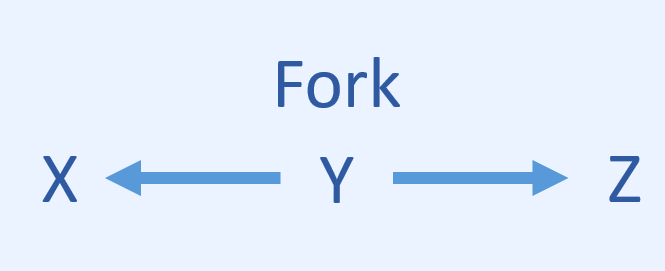
变量除了自己所表示和实际含义外，根据它在路径中的位置以及我们的研究问题，可以分为三类：中介变量，混杂变量，碰撞变量，这也与三种基本因果结构恰好对应。

1. 中介变量与链(chain)结构



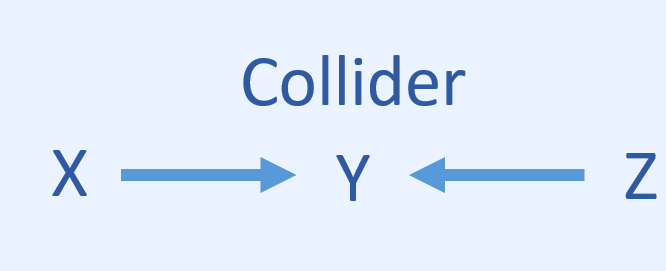
中介变量两端的箭头方向相同。上图X至Z的这条路径中，Y是中介变量。

1. 混杂变量与叉(fork)结构



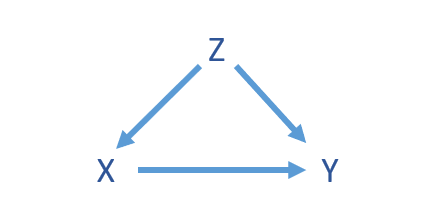
混杂变量两端的箭头尾对尾。上图X至Z的这条路径中，Y是混杂变量。

1. 碰撞变量与对撞(collider)结构



碰撞变量两端的箭头头对头。上图X至Z的这条路径中，Y是碰撞变量。

重申一下，没有哪个变量生来就是中介,混杂or碰撞变量, 取决于我们看问题的视角(视角也可以理解为我们关注的是谁和谁之间的路径, 也就是我们的暴露和结局变量分别是什么)。下面这张图里,有中介,有混杂,还有碰撞变量, 你找到了吗？



## 路径

路径,也即DAG图中从一个变量到另一个变量的一条路。路径是有起点和终点的，默认情况下我们说的就是暴露变量(exposure)E到结局变量(outcome)O之间的路径。所以请别忘记，当我们在谈论路径时，默认的定语就是暴露与结局之间的路径，不会指别的。

### 后门路径 vs 前门路径

后门路径(back-door path) 如果加个定语应该是E的后门路径，就是指这条路径起始于E←。

前门路径(front-door path) 同理，就是指这条路径起始于E→。

### 开放路径 vs 闭塞路径

开放路径(open path, unblocked path) 顾名思义，这条路是通的,也意味着这条路的两个端点的变量应该是有关联的(associated)。

闭塞路径(blocked path) 这条路被阻断了，或者天然就是阻断的。阻断路径意味着切断关联。

"自然"状态下，包含碰撞变量(collider)的路径都是闭塞路径。不含碰撞变量(collider)的路径都是开放路径。

这里的自然是打引号的，是指不控制任何变量的情形下。之所以打引号,是因为有时候我们也没有主动地去控制,但由于无意间选择性地纳入研究人群(经典的例子比如伯克森偏倚Berkson bias),其实是相当于对变量进行了选择,进行了控制,这种情形不属于"自然"状态。

控制一条路径上的碰撞变量(collider)会开放该路径。

控制一条路径上的非碰撞变量(noncollider)会阻断该路径。

以上两条是我们选择变量来控制混杂的依据。

### 直接路径 vs 间接路径

从E到O的路径,如果箭头全都是→,一路向前,就是直接路径。否则就是间接路径。

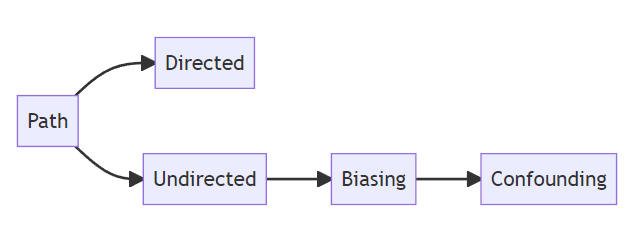
直接路径传递的是因果关联。间接路径传递的是虚假关联。

### 偏倚路径 vs 混杂路径

偏倚路径(biasing path)是指开放的间接路径。天然地(未经控制时)，偏倚路径就是开放的后门路径，反之亦然。

但由于存在人为地或者非人为地控制(例如选择性偏倚)，可能开启本来天然关闭的前门路径，导致出现开放的前门路径，此时也是偏倚路径，但不是混杂路径。

混杂路径(confounding path)是指以→O结尾的偏倚路径。所以，说混杂路径就是开放的后门路径。好像也没毛病呀，有吗？



### 混杂与偏倚控制

控制偏倚的方式是阻断所有偏倚路径。控制混杂的方式是阻断所有混杂路径，也就是阻断所有后门路径。

阻断了所有后门路径意味着没有混杂，但不意味着没有偏倚，可能存在未阻断的前门偏倚路径。

但只要没有人为的控制或者无意间的选择，前门路径中，除了直接路径(即因果效应)外，都是关闭的(你想明白了吗)。也就是自然情形下，给定一个因果图，是不存在开放的前门路径的。

综上，基于因果图,寻找需要控制哪些变量来达到控制混杂目的具体方法就是~

Step1 列举出所有从E到O的后门路径

Step2 找到一个变量集合，确保控制(调整)这些变量后上述所有的后门路径全部被阻断。PS: 这样的集合也可能是空集~ 也可能有多个~