

# 04 图形因果模型：揭示 因果关系

Speaker: 许文立

[wlxu@cityu.edu.mo](mailto:wlxu@cityu.edu.mo)

August-November, 2025

Faculty of Finance, City University of Macau



# 目录

---

因果推断与图形模型

01

---

图形模型的核心结构

02

---

因果推断中的偏误类  
型

03

---

图形模型的应用与总  
结

04

# Part. 01

因果推断与图形模型

---

CHAPTER





## 因果推断的挑战

### 因果推断的复杂性

- 因果推断是一个复杂的过程，其术语往往难以理解。例如，厨师描述食物时，因果关系的表达可能模糊不清。
- 两个人讨论：因为收入的混杂因素，很难识别移民对这个社区的影响，所以我们用了工具变量。



## 因果推断的挑战

### 因果推断的复杂性

图形模型的出现极大地简化了这一过程，它为因果关系提供了一种清晰的语言，能够帮助我们更直观地表达因果机制。

### 图形模型的重要性

- 图形模型是因果推断的重要工具，它能够将复杂的因果关系可视化，帮助我们更好地理解 and 交流。
- 通过图形模型，我们可以清晰地识别因果路径，从而更准确地进行因果推断，避免因术语和逻辑的复杂性而产生的误解。

# 条件独立性的概念



## 条件独立性的定义

条件独立性是因果推断中的一个关键概念，它指的是在给定某些变量的条件下，其他变量之间不再存在依赖关系。例如，在药物治疗中，我们通过分组控制混杂因素，使得处理组和对照组在其他条件上保持一致，从而实现处理与潜在结果的条件独立性。

$$(Y_0, Y_1) \perp T | X$$

## 条件独立性的作用

条件独立性是因果推断的关键假设之一，它有助于我们清晰地识别因果效应。通过控制混杂因素，我们可以排除其他变量对因果关系的干扰，从而更准确地评估处理因素对结果的影响。



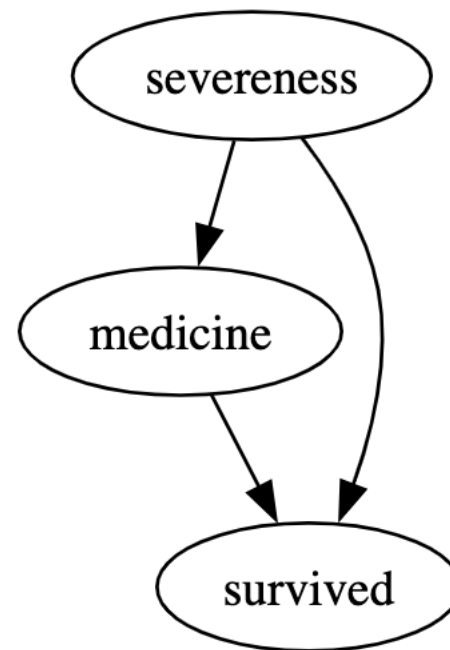
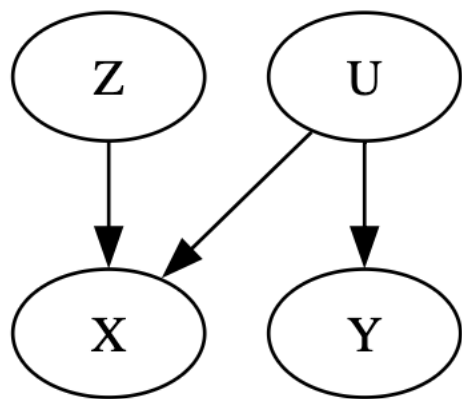
## 实际应用示例

以药物治疗为例，我们通过随机分组的方式，确保治疗组和对照组在年龄、性别等混杂因素上分布相似。这样，**治疗效果与潜在结果**之间的关系就可以在这些混杂因素的控制下被清晰地识别出来，从而实现条件独立性。





## 因果图模型



# Part. 02

图形模型的核心结构

---

CHAPTER

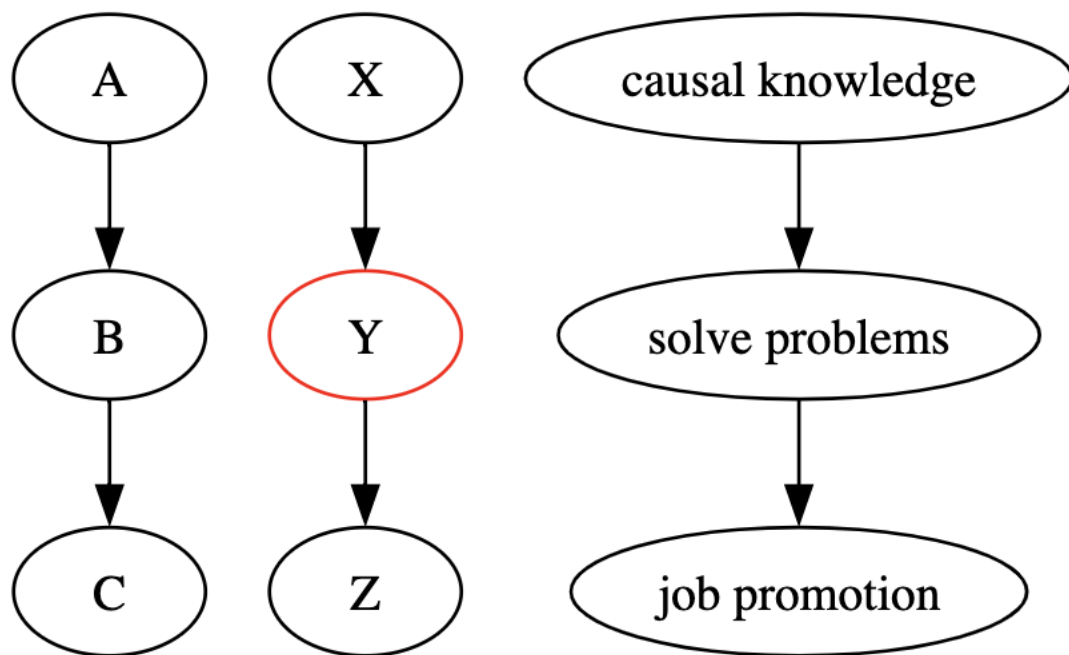


## 链式结构与依赖关系

### 链式结构的特点

在图形模型中，链式结构表现为一个变量导致另一个变量，后者又导致第三个变量。例如，因果知识可以促进问题解决能力的提升，而问题解决能力的提升又可能导致职位晋升。这种结构展示了依赖关系的传递性，即A导致B，B导致C，从而A间接影响C。通过条件化中介变量B，可以阻断A和C之间的依赖关系，实现变量之间的独立性。

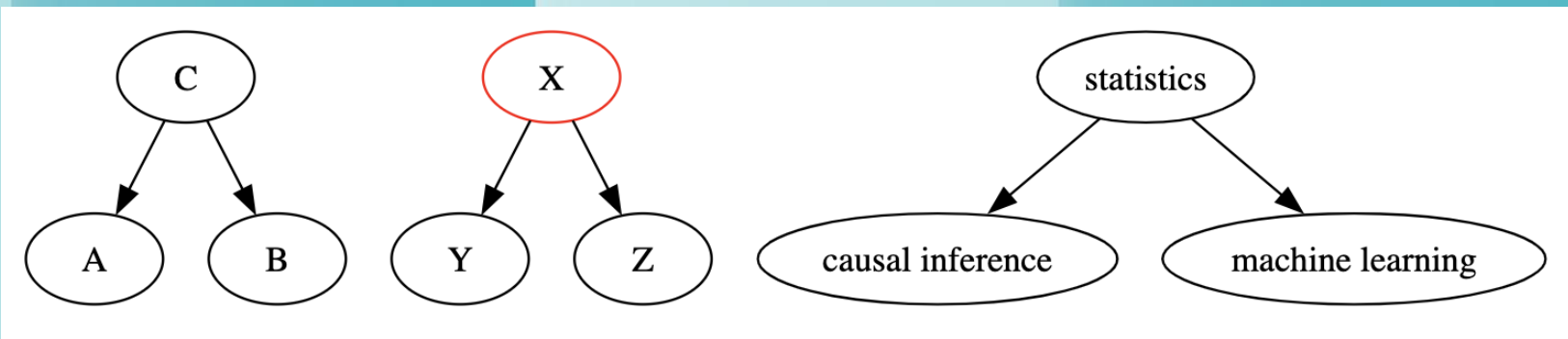
$$A \perp C | B$$



# 分叉结构与后门路径

## 分叉结构的定义

分叉结构是指一个变量同时导致两个其他变量。例如，统计学知识可以同时影响因果推断能力和机器学习能力。这种结构在因果推断中非常常见，它揭示了变量之间的共同原因关系。



**后门路径**是指从处理变量到结果变量的非因果路径，这些路径可能通过共同原因引入混杂偏误。在分叉结构中，如果不加以控制，这些后门路径会导致因果效应的误判。

# 分叉结构与后门路径

## 分叉结构的定义

分叉结构是指一个变量同时导致两个其他变量。例如，统计学知识可以同时影响因果推断能力和机器学习能力。这种结构在因果推断中非常常见，它揭示了变量之间的共同原因关系。

01

## 后门路径的概念

后门路径是指从处理变量到结果变量的非因果路径，这些路径可能通过共同原因引入混杂偏误。在分叉结构中，如果不加以控制，这些后门路径会导致因果效应的误判。

02

## 阻断后门路径的方法

为了阻断后门路径，我们需要条件化共同原因。例如，在研究教育对收入的影响时，家庭背景是一个共同原因。通过控制家庭背景，我们可以阻断教育和收入之间的后门路径，从而更准确地估计教育对收入的因果效应。

$$A \perp B | C$$

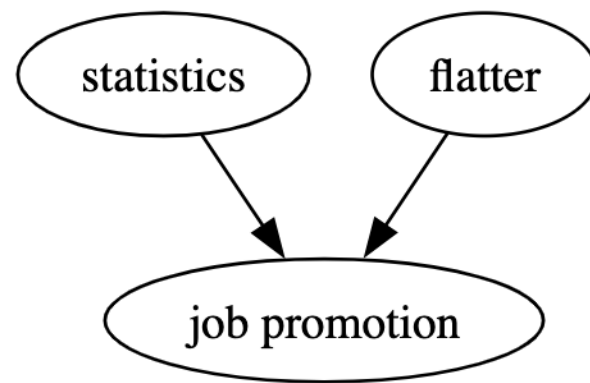
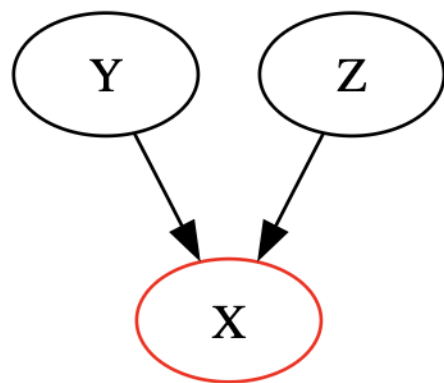
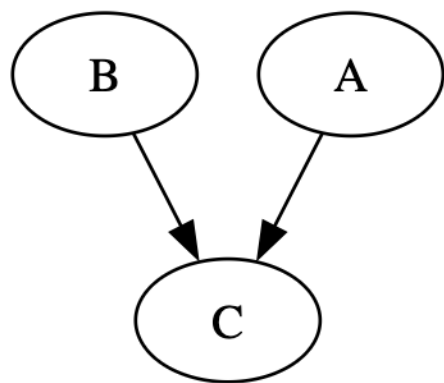
03

## 实际应用的重要性

分叉结构和后门路径的识别与处理是因果推断中的关键环节。通过图形模型，我们可以清晰地识别这些结构，并采取合适的控制方法，从而避免混杂偏误，提高因果推断的准确性。

04





## 对撞节点的定义

对撞节点是指两个变量共享一个共同的影响结果。例如，统计学知识和奉承上司都可能影响职位晋升。在这种结构中，晋升是统计学知识和奉承上司的对撞节点。

## 对撞节点与解释消除



## 对撞节点的定义

对撞节点是指两个变量共享一个共同的影响结果。例如，统计学知识和奉承上司都可能影响职位晋升。在这种结构中，晋升是统计学知识和奉承上司的对撞节点。

## 条件化对撞节点的影响

对撞节点的条件化会打开依赖路径。例如，如果我们直接控制晋升这一结果变量，就会引入统计学知识和奉承上司之间的依赖关系，从而导致偏误。而不条件化对撞节点，则可以保持路径关闭，避免这种偏误。

$$A \not\perp B|C$$

## 实际应用中的注意事项

在实际因果推断中，对撞节点的存在可能会导致意外的偏误。因此，在使用图形模型时，我们需要仔细识别对撞节点，并避免对其进行不恰当的条件化，以确保因果推断的准确性。

# 对撞节点与解释消除

「LOGO」

# Part. 03

## 因果推断中的偏误类型

了解因果图模型能让我们认识到因果推断中出现的问题。正如所见，这些问题归根结底都源于偏误。

$$E[Y|T = 1] - E[Y|T = 0] = \underbrace{E[Y_1 - Y_0|T = 1]}_{ATE} + \underbrace{\{E[Y_0|T = 1] - E[Y_0|T = 0]\}}_{BIAS}$$

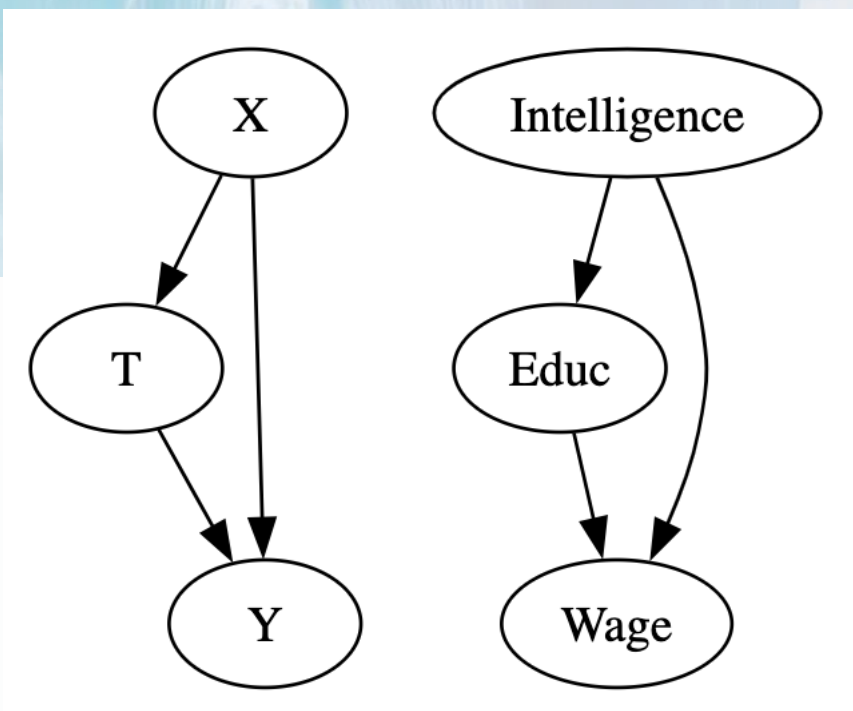
图形模型帮助我们诊断所面临的偏误类型，并确定纠正这些偏误所需的工具。



# 混杂偏误的产生与解决

## 混杂偏误的定义

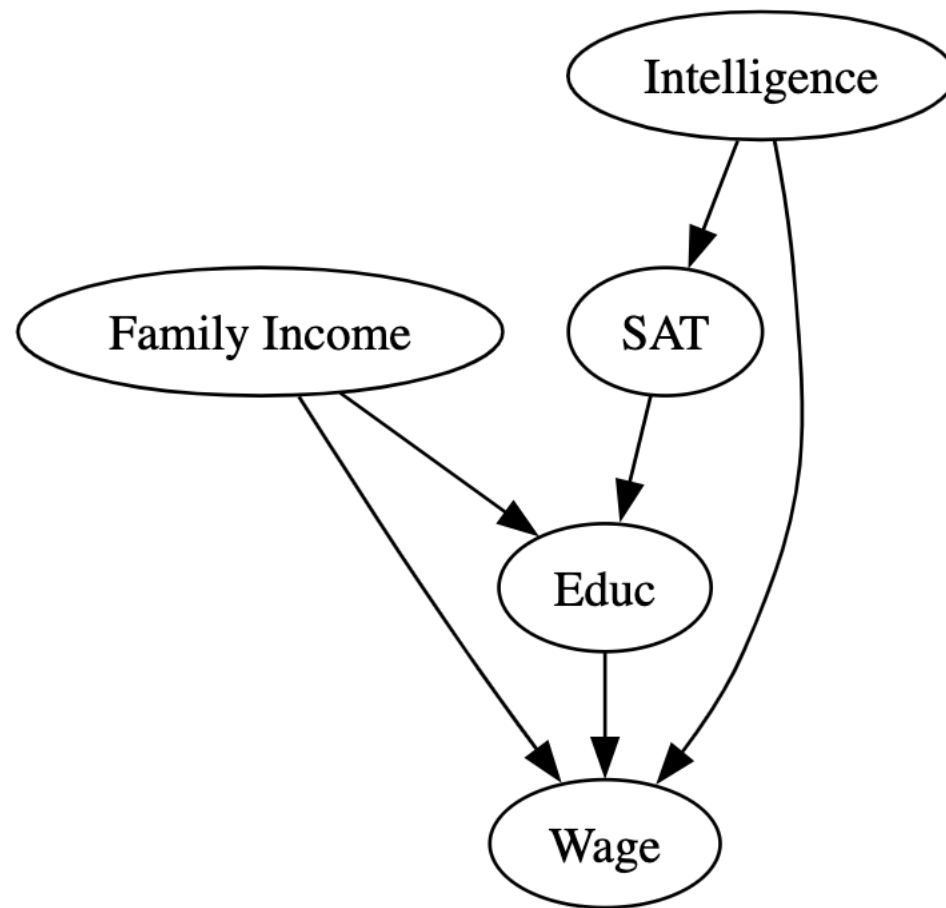
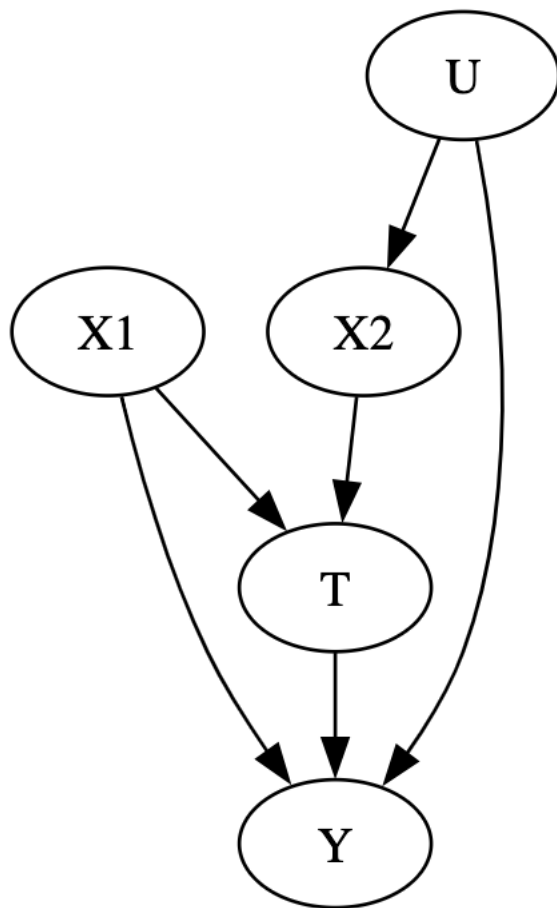
混杂偏误是指处理因素与结果之间存在共同原因时产生的偏误。例如，教育程度和收入水平之间可能存在混杂偏误，因为家庭背景可能同时影响教育机会和收入水平。这种偏误会干扰我们对因果效应的准确估计。



# 混杂偏误的产生与解决

## 混杂偏误的产生与解决

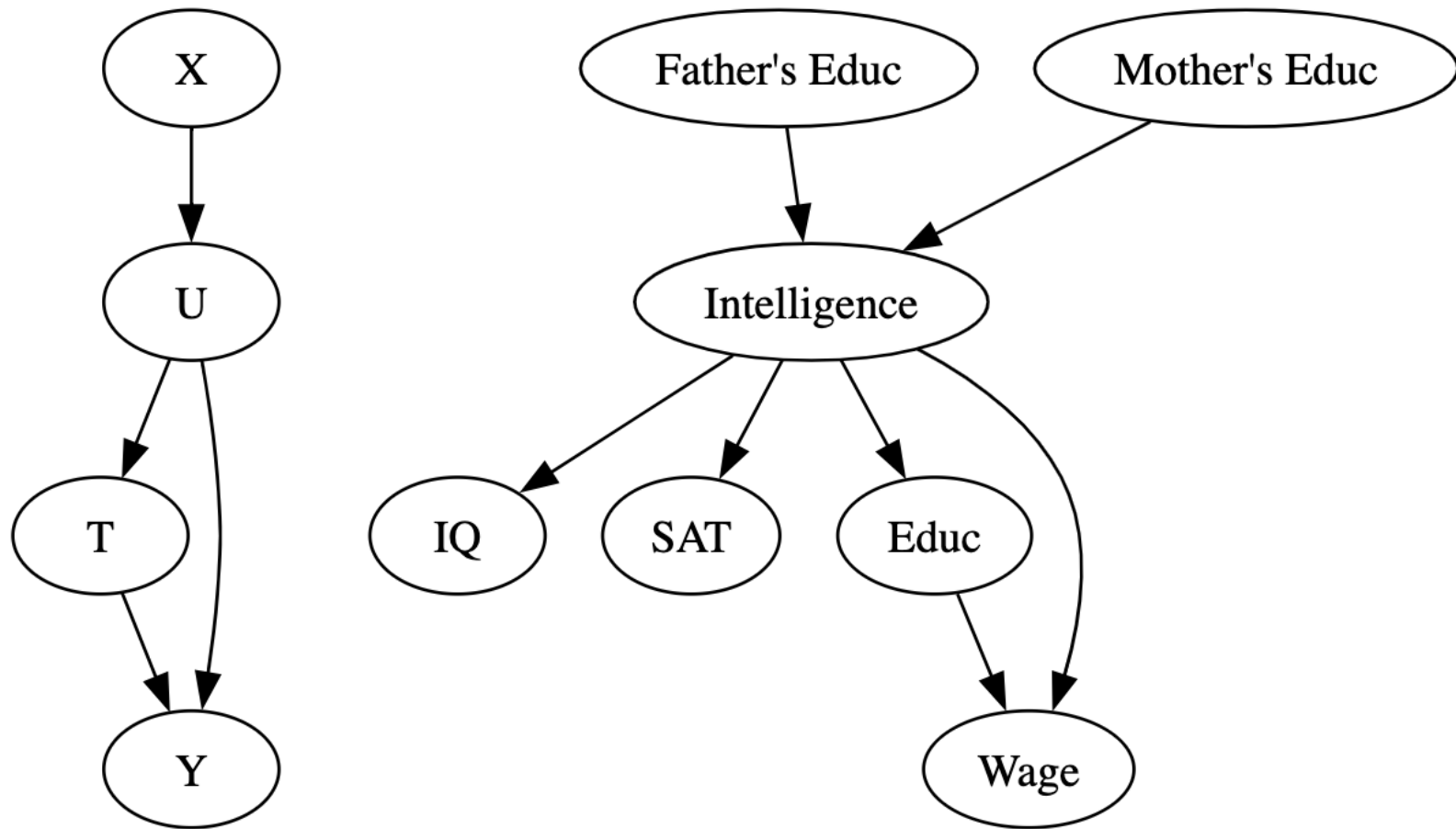
为了纠正混杂偏误，我们需要控制共同原因。在无法直接控制所有共同原因时，可以使用替代混杂因子。例如，通过控制家庭收入等替代变量，我们可以阻断教育和收入之间的后门路径，从而更准确地估计教育对收入的因果效应。



# 混杂偏误的产生与解决

## 混杂偏误的产生与解决

为了纠正混杂偏误，我们需要控制共同原因。在无法直接控制所有共同原因时，可以使用替代混杂因子。例如，通过控制家庭收入等替代变量，我们可以阻断教育和收入之间的后门路径，从而更准确地估计教育对收入的因果效应。





# 选择偏误的来源与影响

## 选择偏误的定义与影响

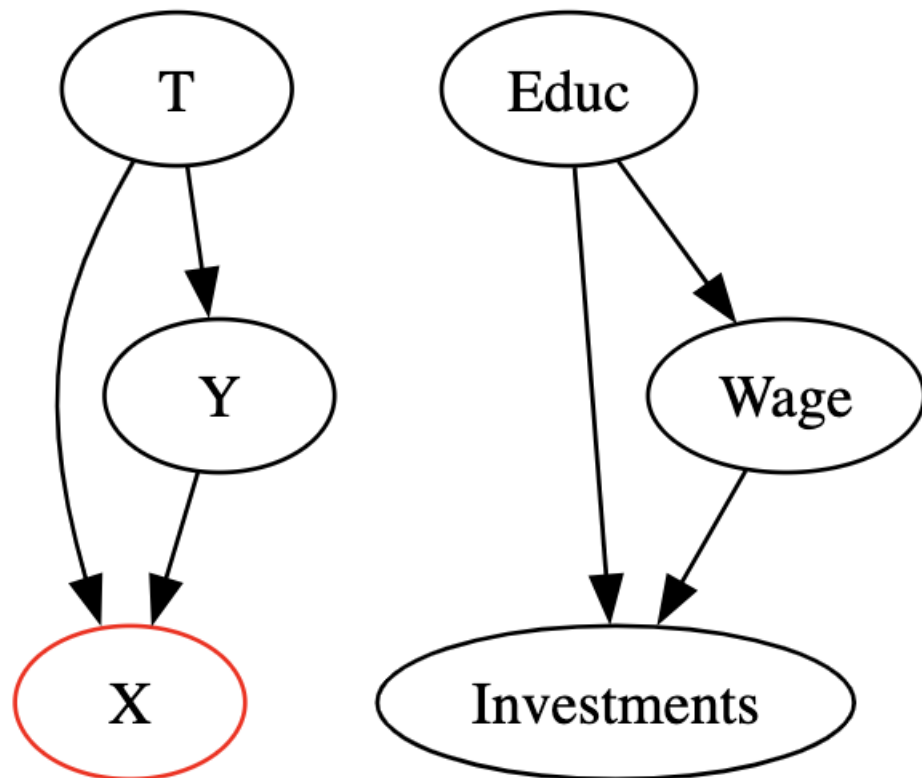
选择偏误是指在因果推断中，控制过多变量导致的偏误。例如，在研究教育对工资的影响时，如果控制了对撞变量或中介变量，如职业选择，可能会引入偏误。选择偏误的危险性在于过度控制可能导致因果效应估计的偏差，从而误导研究结论。



Moving on to the next point



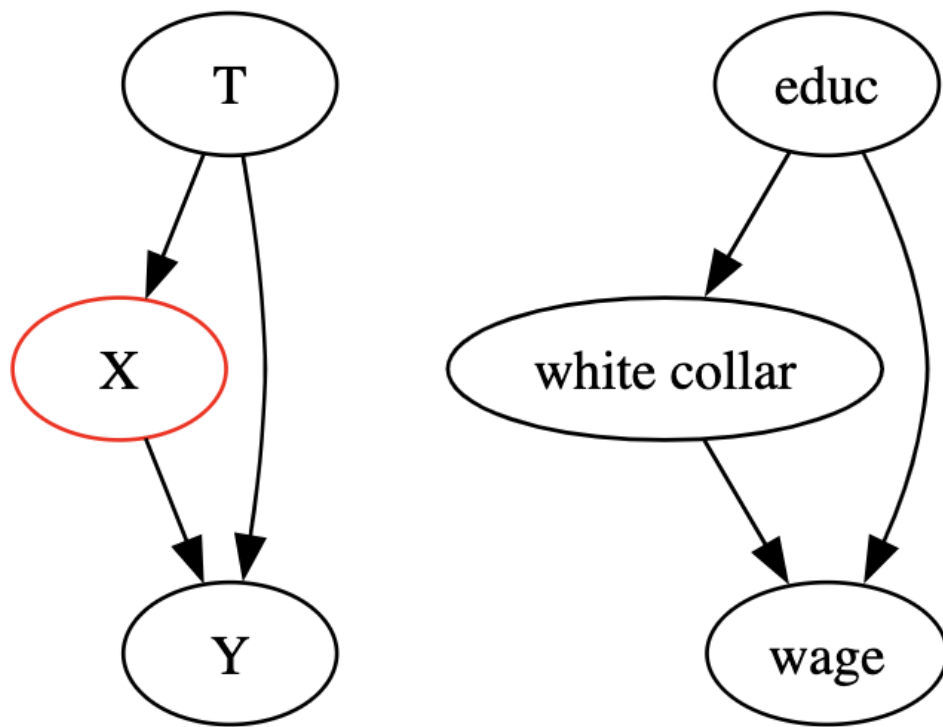
## 选择偏误的来源与影响



Moving on to the next point



## 选择偏误的来源与影响（中介变量）



Moving on to the next point





# Part. 04

图形模型的应用与总结

---

CHAPTER

# 图形模型的 诊断能力

## 1 图形模型的作用

图形模型在因果推断中具有强大的诊断能力。它能够帮助我们识别因果结构中的潜在问题，如混杂偏误和选择偏误。通过清晰地表示变量之间的关系，图形模型为研究者提供了一个直观的工具，用于分析和纠正偏误。

## 2 确定纠正偏误的工具

图形模型不仅能够识别偏误，还能帮助我们确定纠正偏误所需的工具。例如，通过识别后门路径，我们知道需要控制哪些变量来阻断这些路径，从而实现条件独立性。这种能力使得图形模型在实际应用中具有极高的价值。

## 3 实际应用示例

在研究广告投放对销售的影响时，图形模型可以帮助我们识别广告投放与销售之间的因果路径，以及可能存在的混杂因素。通过控制这些混杂因素，我们可以更准确地估计广告投放的因果效应，从而为营销决策提供科学依据。

# 核心要点回顾

## 条件独立性的规则

条件独立性是因果推断中的关键规则，它要求在给定某些变量的条件下，其他变量之间不再存在依赖关系。通过控制混杂因素，我们可以实现条件独立性，从而更准确地识别因果效应。

## 混杂偏误的识别与纠正

混杂偏误是指处理因素与结果之间存在共同原因时产生的偏误。通过控制共同原因，我们可以阻断后门路径，从而纠正混杂偏误。在无法直接控制所有共同原因时，可以使用替代混杂因子。

## 选择偏误的危险性

选择偏误是指在因果推断中，控制过多变量导致的偏误。例如，控制对撞变量或中介变量可能会引入偏误。选择偏误的危险性在于过度控制可能导致因果效应估计的偏差，从而误导研究结论。

## 图形模型的重要性

图形模型是理解和表达因果关系的重要工具。它能够帮助我们识别和纠正偏误，通过清晰表示因果结构，为研究者提供了一个直观的分析框架，从而实现更准确的因果推断。



谢谢  
Thank you for listening

