

独立性 VS 不相关

Independence versus Zero Correlation

高中统计学 · 第肆课



何濯羽



2024 年 2 月 15 日

目录

CONTENTS



01 联合分布

Joint Distribution



02 统计独立性

Statistical Independence



03 条件期望 & 均值独立

Conditional Expectation & Mean Independence



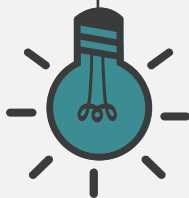
04 线性不相关

Uncorrelatedness

01

联合分布

Joint Distribution





联合分布的定义

在第貳课和第叁课中，我们只考虑了单个随机变量的概率分布和矩。本节课中，我们将开始研究两个（离散型）随机变量共同的概率分布以及它们之间的特殊关系。

两个变量共同的概率分布被称为“**联合分布**”（joint distribution）。对于两个离散型随机变量，它们所有可能的取值以及相应的概率可以由一张“联合分布列”展示出来。

		Y				
		y_1	y_2	\cdots	y_m	总和
X	x_1	p_{11}	p_{12}	\cdots	p_{1m}	$p_{1\cdot}$
	x_2	p_{21}	p_{22}	\cdots	p_{2m}	$p_{2\cdot}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	x_n	p_{n1}	p_{n2}	\cdots	p_{nm}	$p_{n\cdot}$
总和		$p_{\cdot 1}$	$p_{\cdot 2}$	\cdots	$p_{\cdot m}$	



联合分布的CDF与PMF

$F_{X,Y}(x, y) = \Pr(X \leq x, Y \leq y)$ 被称为“联合累积分布函数” (joint CDF) 。

$f_{X,Y}(x, y) = \Pr(X = x, Y = y)$ 被称为“联合概率质量函数” (joint PMF) 。

例:

例:

		Y				X 的边际分布
		0	2	4	6	总和
X	0	0.1	0.05	0.08	0.02	0.25
	1	0.15	0.1	0.05	0.05	0.35
	2	0.13	0.1	0	0.07	0.3
	3	0.02	0.05	0.02	0.01	0.1
总和		0.4	0.3	0.15	0.15	1

Y 的边际分布

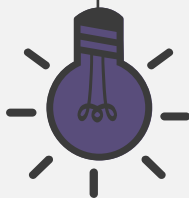
箭

$$\begin{aligned} F_{X,Y}(2,2) &= \Pr(X \leq 2, Y \leq 2) \\ &= \Pr(X = 0, Y = 0) + \Pr(X = 0, Y = 2) \\ &\quad + \Pr(X = 1, Y = 0) + \Pr(X = 1, Y = 2) \\ &\quad + \Pr(X = 2, Y = 0) + \Pr(X = 2, Y = 2) \\ &= 0.1 + 0.05 + 0.15 + 0.1 + 0.13 + 0.1 \\ &= 0.65 \end{aligned}$$

符合概率公理的要求

02 统计独立性

Statistical Independence





定义统计独立性

如果事件 A 和 事件 B 满足

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

那么我们称 A 和 B 为**独立事件** (independent events) 。

如果离散型随机变量 X 和 Y 满足

$$\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y), \text{ 即 } f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y), \quad \forall x, y \in \mathbb{R}$$

或

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y), \quad \forall x, y \in \mathbb{R}$$

那么我们称 X 和 Y 相互**统计独立** (statistically independent) , 或相互**随机独立** (stochastically independent) , 简称 “相互独立” 。



例：独立性的应用

已知随机变量 X 和 Y 相互独立，你能根据它们的边际分布填满整张联合分布列吗？

		Y			
		0	1	2	总和
X	0	0.08	0.12	0.2	0.4
	1	0.02	0.03	0.05	0.1
	2	0.1	0.15	0.25	0.5
总和		0.2	0.3	0.5	

原理：独立 $\Rightarrow \Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$



强大的独立性

如果事件 A 和 B **相互独立**, 那么 $\Pr(A|B) = \Pr(A)$ 且 $\Pr(B|A) = \Pr(B)$ 。

如果随机变量 X 和 Y **相互独立**, 那么 $\Pr(X = x|Y = y) = \Pr(X = x)$ 且 $\Pr(Y = y|X = x) = \Pr(Y = y)$, $\forall x, y$ 。

如果随机变量 X 和 Y **相互独立** 且 $g(\cdot)$ 和 $h(\cdot)$ 是两个函数, 那么 $g(X)$ 和 $h(Y)$ 也相互独立。

如果随机变量 X 和 Y **相互独立** 且存在两个函数 $g: \mathbb{R} \rightarrow \mathbb{R}$ 和 $h: \mathbb{R} \rightarrow \mathbb{R}$, 那么

$$E(XY) = E(X) \cdot E(Y)$$

且

$$E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)]$$

注意: 一般情况下, $E(XY) \neq E(X) \cdot E(Y)$ 。



独立 VS 互斥

事件 A 和 B **相互独立** $\Leftrightarrow \Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$

事件 A 和 B **互斥** $\Leftrightarrow A \cap B = \emptyset$

从数学定义上看，“独立”和“互斥”是两个完全不同的概念。

如果 $\Pr(A) > 0$ 且 $\Pr(B) > 0$ ，那么

- 事件 A 和 B 相互独立 \Rightarrow 事件 A 和 B 不互斥
- 事件 A 和 B 互斥 \Rightarrow 事件 A 和 B 不相互独立



独立 VS 互斥

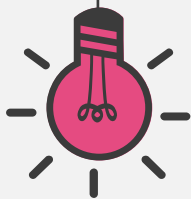
证明:

- 当 A 和 B 相互独立时, 则必有 $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$ 。【反证法】假设 A 和 B 互斥, 则 $A \cap B = \emptyset$, 这意味着 $\Pr(A \cap B) = \Pr(\emptyset) = 0$ 。这与 $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B) > 0$ 相悖, 因此, 假设错误, 即 A 和 B 不互斥。
- 当 A 和 B 互斥时, 则必有 $A \cap B = \emptyset$ 。因为 $\Pr(A \cap B) = \Pr(\emptyset) = 0$ 和 $\Pr(A) \cdot \Pr(B) > 0$, 所以 $\Pr(A \cap B) \neq \Pr(A) \cdot \Pr(B)$, 即 A 和 B 不相互独立。

03

条件期望 & 均值独立

Conditional Expectation & Mean Independence



注：这一节的内容不属于浙江数学高考范围，也不在人教版数学教材中，但是我认为这些知识点是不可忽略的。这方面知识的缺失，不利于之后线性回归模型的教学。

实际上，这一节的内容是统计学中大多数线性/非线性回归模型的基石。



条件分布 (Conditional Distribution)

请回忆一下我们在第壹课中学习的**条件概率定义式**：

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

相似地，我们可以定义条件概率质量函数 (conditional PMF) 和条件累积分布函数 (conditional CDF)：

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\Pr(X=x, Y=y)}{\Pr(X=x)} \text{ 和 } F_{Y|X}(y|x) = \frac{\Pr(Y \leq y, X=x)}{\Pr(X=x)}$$

其中， X 和 Y 均为离散型随机变量。



条件期望的定义

假设随机变量 Y 在给定 $X = x$ 时的条件 PMF 为 $f_{Y|X}(y|x)$ 。 Y 在给定 $X = x$ 时的**条件期望** (conditional expectation) 是

$$E(Y|X = x) = \sum_{y \in \mathcal{Y}} y \cdot f_{Y|X}(y|x) = \sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X = x) = \sum_{y \in \mathcal{Y}} y \cdot \frac{\Pr(Y = y, X = x)}{\Pr(X = x)}$$

重点：数学期望 $E(Y)$ 是一个定值，而条件期望 $E(Y|X = x)$ 是一个以 x 为自变量的函数！



例：条件期望的计算

已知随机变量 X 和 Y 拥有如下的联合分布列，请计算 $E(X|Y=2)$ 。

		X			
		1	2	3	总和
Y	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$
	3	$\frac{1}{3}$	0	$\frac{1}{6}$	$\frac{1}{2}$
	4	0	$\frac{1}{6}$	0	$\frac{1}{6}$
总和		$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	

先计算条件概率：

$$\Pr(X=1|Y=2) = \frac{\Pr(X=1, Y=2)}{\Pr(Y=2)} = \frac{\frac{1}{12}}{\frac{1}{3}} = \frac{1}{4}$$

$$\Pr(X=2|Y=2) = \frac{\Pr(X=2, Y=2)}{\Pr(Y=2)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}$$

$$\Pr(X=3|Y=2) = \frac{\Pr(X=3, Y=2)}{\Pr(Y=2)} = \frac{\frac{1}{12}}{\frac{1}{3}} = \frac{1}{4}$$

再计算条件期望：

$$E(X|Y=2) = \sum_{x=1}^3 x \cdot \Pr(X=x|Y=2) = 1 \times \frac{1}{4} + 2 \times \frac{1}{2} + 3 \times \frac{1}{4} = 2$$



例：条件期望的计算

已知随机变量 X 和 Y 拥有如下的联合分布列，请计算 $E(X|Y=2)$ 。

		X			
		1	2	3	总和
Y	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$
	3	$\frac{1}{3}$	0	$\frac{1}{6}$	$\frac{1}{2}$
	4	0	$\frac{1}{6}$	0	$\frac{1}{6}$
总和		$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	

相似地，我们可以算出所有的 $E(X|Y)$ ：

$$E(X|Y=2) = 2, \quad E(X|Y=3) = \frac{5}{3}, \quad E(X|Y=4) = 2$$

通过这个例题，我们可以清楚地看见条件期望 $E(X|Y)$ 是一个以随机变量 Y 为自变量的函数，记为 $m(Y)$ 。当 Y 的取值 y 给定时， $m(Y)$ 会返回一个定值 $m(y)$ 。

在此例中，

$$m(y) = E(X|Y=y) = \begin{cases} 2 & (y=2) \\ \frac{5}{3} & (y=3) \\ 2 & (y=4) \end{cases}$$



期望迭代法则

如果 $E(|Y|) < \infty$, 那么对于任意随机变量 X , 以下等式恒成立:

$$E[E(Y|X)] = E(Y)$$

证明: $E(Y|X) = \sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X)$

这个法则在许多统计学的证明中大放异彩。

$$\begin{aligned} E[E(Y|X)] &= \sum_{x \in \mathcal{X}} \left\{ \left[\sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X = x) \right] \cdot \Pr(X = x) \right\} \\ &= \sum_{x \in \mathcal{X}} \left[\sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X = x) \cdot \Pr(X = x) \right] \\ &= \sum_{x \in \mathcal{X}} \left[\sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y, X = x) \right] \\ &= \sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y) = E(y) \end{aligned}$$



均值独立

当且仅当 $E(Y|X = x) = E(Y) \forall x$, 我们称随机变量 Y 均值独立 (mean independent) 于随机变量 X 。

当 Y 与 X 相互统计独立时, Y **必定** 也均值独立于随机变量 X 。

证:
$$E(Y|X = x) = \sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X = x) = \sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y) = E(Y)$$

“统计独立” 保证了该等号的成立。



均值独立

当 Y 均值独立于随机变量 X 时, Y 与 X **不一定**相互统计独立。

例:

		X			
		-1	0	1	总和
Y	-1	0.15	0	0.15	0.3
	0	0	0.4	0	0.4
	1	0.15	0	0.15	0.3
总和		0.3	0.4	0.3	

计算可得

$$E(Y|X) = E(Y) = 0$$

但是,

$$\exists x, y, \Pr(X = x, Y = y) \neq \Pr(X = x) \cdot \Pr(Y = y)$$

因此, Y 均值独立于 X , 但 Y 与 X 不相互统计独立。



均值独立

均值独立**不具有对称性**: X 均值独立于 $Y \not\Leftrightarrow Y$ 均值独立于 X 。

例:

		X			
		-1	0	1	总和
Y	-1	0.15	0	0.1	0.25
	0	0	0.5	0	0.5
	1	0.15	0	0.1	0.25
总和		0.3	0.5	0.2	

计算可得

$$E(Y|X) = E(Y) = 0$$

但是,

$$E(X|Y = y) = \begin{cases} -0.2 & (y = -1) \\ 0 & (y = 0) \\ -0.2 & (y = 1) \end{cases} \neq E(X)$$

因此, Y 均值独立于 X , 但 X 并不均值独立于 Y 。

04 线性不相关

Uncorrelatedness





协方差的定义与作用

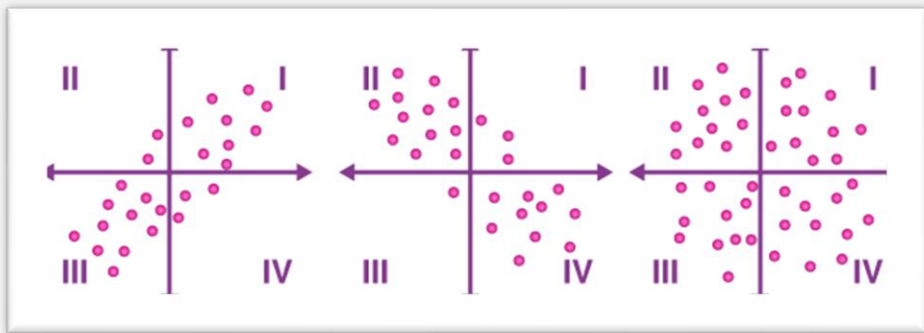
随机变量 X 与 Y 的**协方差** (covariance) 是

$$\sigma_{X,Y} = Cov(X,Y) = E\{[X - E(X)] \cdot [Y - E(Y)]\} = E(XY) - E(X) \cdot E(Y)$$

我们可以把 $Cov(X,Y)$ 理解为 X 与 Y 所服从的**联合分布的二阶中心矩**。

协方差可以衡量两个随机变量的线性相关程度 (linear relatedness) :

- $Cov(X,Y) > 0$ 意味着 X 与 Y 线性**正相关**。
- $Cov(X,Y) < 0$ 意味着 X 与 Y 线性**负相关**。
- $Cov(X,Y) = 0$ 意味着 X 与 Y 线性**不相关**。





协方差的性质

协方差 $Cov(X, Y)$ 具有以下性质:

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, a) = 0$
- $Cov(aX \pm bY, cZ) = ac \cdot Cov(X, Z) \pm bc \cdot Cov(Y, Z)$
- $D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$

其中, 大写字母代表随机变量, 小写字母代表常数。

练习: 证明这些性质的成立。



协方差的缺陷

协方差的范围是 $(-\infty, +\infty)$ 。可是，协方差具有与方差相同的问题：因为随机变量的单位变化势必导致协方差的大小变化，所以协方差的数值大小并不能有效反映线性相关程度的大小。

为了化解单位变量带来的影响，我们创造了“标准差”以对应方差。相似地，我们也可以创造新的统计工具以对应协方差——“相关系数”！



相关系数的定义

随机变量 X 与 Y 的**相关系数** (correlation coefficient) 是

$$\rho_{X,Y} = \text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

我们可以把 $\text{Corr}(X,Y)$ 理解为向量 $\bar{\mathbf{X}}$ 与 $\bar{\mathbf{Y}}$ (中心化随机变量) 夹角 θ 的余弦值。

$$\text{Corr}(X,Y) = \frac{\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} (x - \mu_X) \cdot (y - \mu_Y) \cdot \Pr(X = x, Y = y)}{\sqrt{\sum_{x \in \mathcal{X}} (x - \mu_X)^2 \cdot \Pr(X = x)} \cdot \sqrt{\sum_{y \in \mathcal{Y}} (y - \mu_Y)^2 \cdot \Pr(Y = y)}}$$

$$\cos \theta = \frac{\bar{\mathbf{X}} \cdot \bar{\mathbf{Y}}}{\|\bar{\mathbf{X}}\| \cdot \|\bar{\mathbf{Y}}\|} = \frac{(\mathbf{X} - \boldsymbol{\mu}_X) \cdot (\mathbf{Y} - \boldsymbol{\mu}_Y)}{\|\mathbf{X} - \boldsymbol{\mu}_X\| \cdot \|\mathbf{Y} - \boldsymbol{\mu}_Y\|}$$

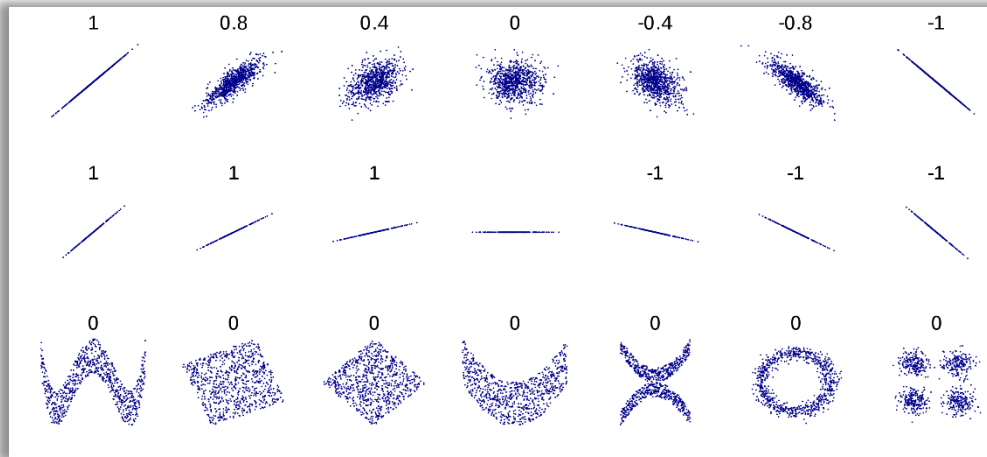


相关系数的含义

该结论的证明需要使用柯西-施瓦茨不等式，在此不做展示。

相关系数 $\text{Corr}(X, Y)$ 的范围是 $[-1, 1]$ ，其含义为

- $\text{Corr}(X, Y) = 1$ 意味着 X 与 Y **完美线性正相关**。
- $\text{Corr}(X, Y) \in (0, 1)$ 意味着 X 与 Y **线性正相关**。
- $\text{Corr}(X, Y) = 0$ 意味着 X 与 Y **线性不相关**。
- $\text{Corr}(X, Y) \in (-1, 0)$ 意味着 X 与 Y **线性负相关**。
- $\text{Corr}(X, Y) = -1$ 意味着 X 与 Y **完美线性负相关**。





线性不相关

如果 $Cov(X, Y) = 0$ 或 $Corr(X, Y) = 0$ ，那么我们称随机变量 X 与 Y **线性独立** (linear independent) 或**线性不相关** (uncorrelated)。

当 Y 均值独立于 X 时， Y **必定**与 X 线性不相关。

证： $Cov(X, Y) = E\{[X - E(X)] \cdot [Y - E(Y)]\} \stackrel{①}{=} E(E\{[X - E(X)] \cdot [Y - E(Y)]|X\}) = E\{[X - E(X)] \cdot E[Y - E(Y)|X]\}$

因为 $E[Y - E(Y)|X] = E(Y|X) - E[E(Y)|X] \stackrel{②}{=} E(Y|X) - E(Y) = E(Y) - E(Y) = 0$,

所以 $Cov(X, Y) = E\{[X - E(X)] \cdot E[Y - E(Y)|X]\} = E\{[X - E(X)] \cdot 0\} = 0$ 。

① 期望迭代法则

② 均值独立



线性不相关

当 Y 与 X 线性不相关时, Y **不一定** 均值独立于 X 。

例:

		X			
		-1	0	1	总和
Y	0	0.1	0.4	0.1	0.6
	1	0.2	0	0.2	0.4
总和		0.3	0.4	0.3	

计算可得

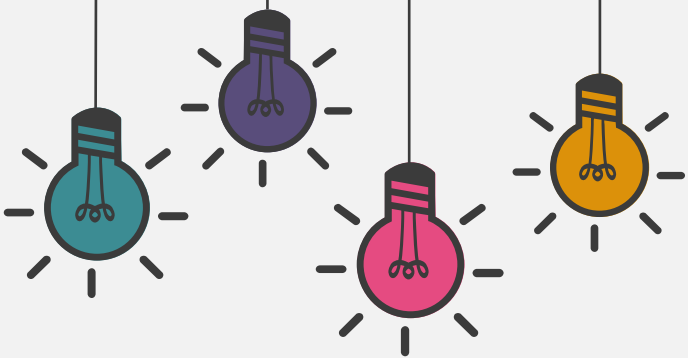
$$E(X) = 0, \quad E(Y) = 0.4$$

$$\text{Cov}(X, Y) = E[(X - 0)(Y - 0.4)] = 0$$

但是,

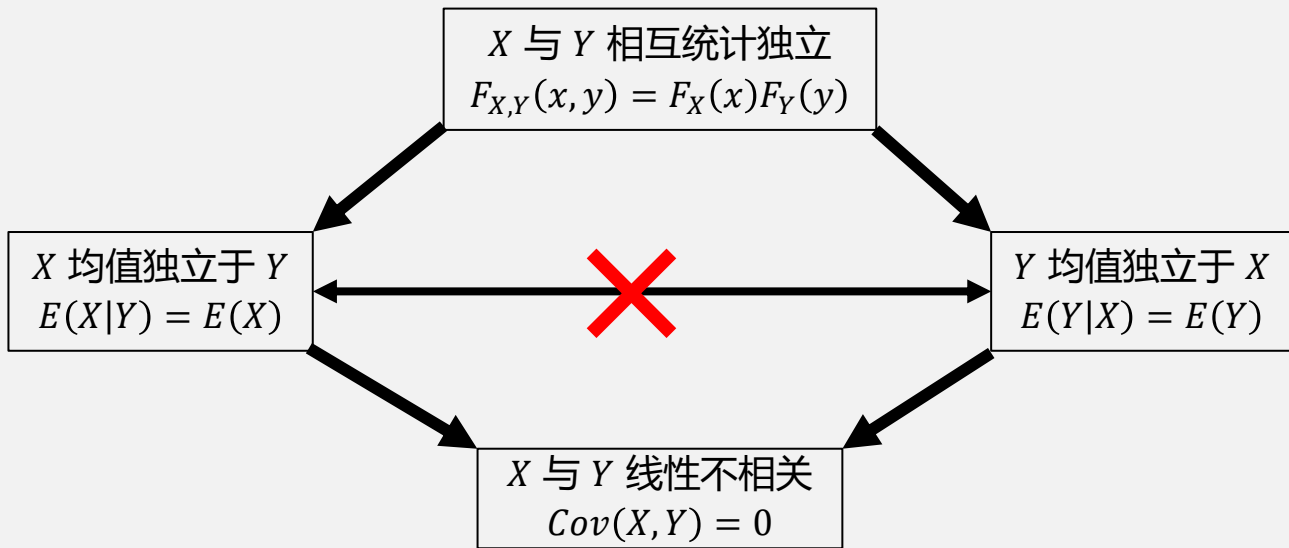
$$E(Y|X = x) = \begin{cases} \frac{2}{3} & (x = -1) \\ 0 & (x = 0) \\ \frac{2}{3} & (x = 1) \end{cases} \neq E(Y)$$

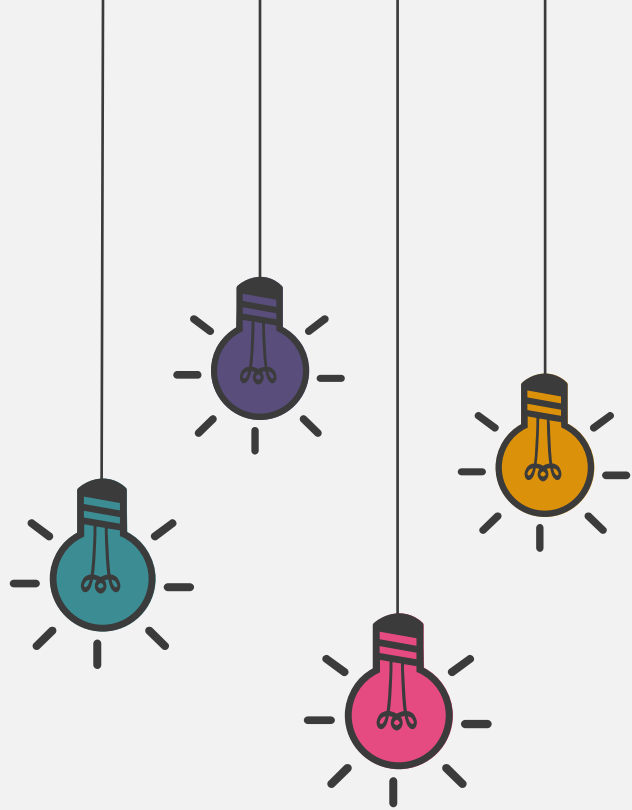
因此, Y 与 X 线性不相关, 但 Y 并不均值独立于 X 。



总结

统计独立（简称“独立”）是一个非常强劲的变量关系，强于均值独立，强于线性不相关。许多统计学的定理都建立在“独立”假设上。可是在实际研究中，独立是难以严格证明的。许多研究者退而求其次，只证明均值独立，用以说明独立“可能”成立。





感谢您的聆听!

Thanks for your listening!

高中统计学·第肆课



何濯羽



2024年2月15日