2014.4.1

第十三讲. 多重线性回归

1. 案例：性别歧视诉讼案
2. 多重线性回归模型
3. 最小二乘法

1. 案例：性别歧视诉讼案

数据集 ***salary (alr3)*** 是美国中西部一个大学在**80**年代关于**“**女性工资待遇受歧视**”**的法律诉讼过程中出示的数据。该数据中所有**52**人都是学校正式教工。

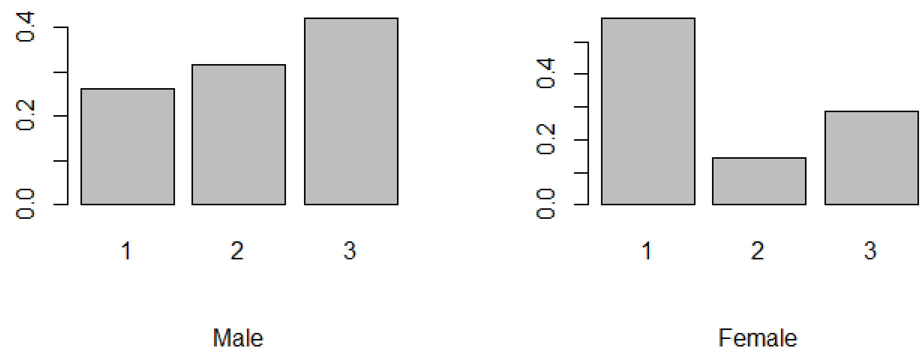
数据表明女性平均工资比男性低**3340$**

假设工资**(Salary)**的对数服从正态，两样本**t-**检验得**p**值**p=0.048,** 在**0.05**水平下可以认为男女工资有差异**.**

或者等价地**,**在如下简单线性模型中检验 $H_0: b = 0$

$$\log(Salary) = a + b \times Sex + \varepsilon,\ \ \varepsilon \sim (0, \sigma^2)$$

但这种差异可能是其它因素引起的, 比如: 男女两组的职称等是否有系统性差异？ (如果男性职称普遍较高, 因为职称越高工资越高, 则可能导致男性工资偏高).



干扰因素包括职称,学历,工龄, 在现职称上的年数 (下表).

变量	描述
Sex	**1** : 女， **0**：男
Rank	职称 **1: Assistant Prof, 2: Associate Prof, 3:Full Prof**
Year	拥有当前职称（**rank**）的时间（单位：年）
Degree	最高学位。 **1**：博士, **0**：硕士
YSdeg	工龄: 获得最高学位至今的时间（单位：年）
Salary	年薪 **($)**

如何控制这些干扰因素/协变量?

在简单线性模型的右端添加若干项:

$$\log(Salary) = a + b \times Sex + c \times Rank + d \times YS\,\mathrm{deg}$$
$$+ e \times Degree + f \times Year + \varepsilon$$

即可达到控制Rank, YEdeg, Degree, Year的目的.

检验H0: b=0,　p值p=0.26,不显著，即男女工资无差异**.**

问题是:

■ 在简单模型右端加上若干项（即多重回归）为什么就达到了控制变量的目的?

(下页)

■ 线性可加模型实际上假定了不同的**Rank**等级内男女工资差异**(b)**是一致的（对其它变量也是这样），这是否与实际相符?

(以后我们将使用交互作用检验线性可加的合理性**)**

简单模型：

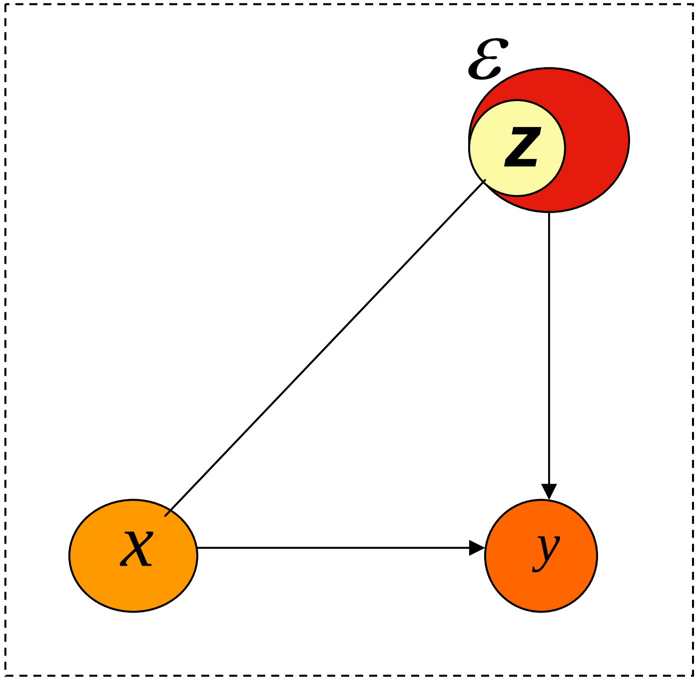
$y = a + bx + \varepsilon, \quad \varepsilon \sim (0, \sigma^2),$

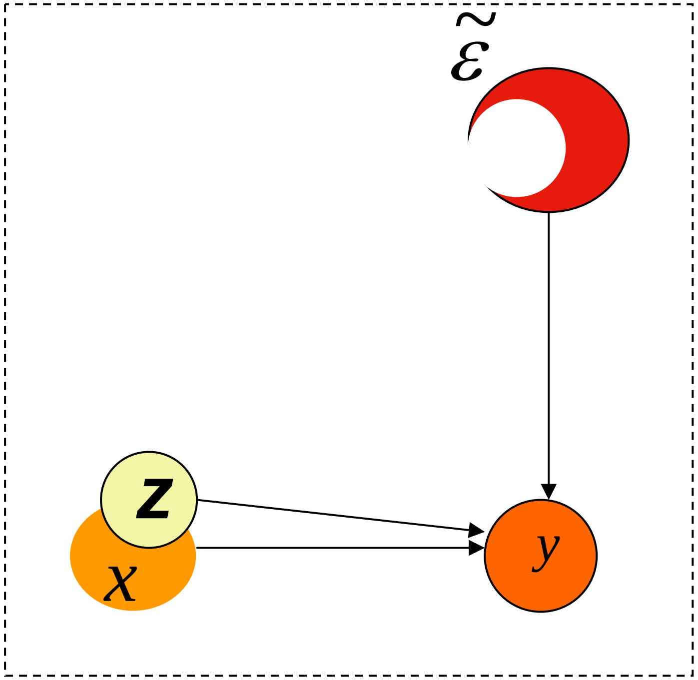
ε中含与x有关的成分

多重回归模型

$y = a + bx + cz + \tilde{\varepsilon}, \quad \tilde{\varepsilon} \sim (0, \tau^2),$

$\tilde{\varepsilon}$与(x, z)独立





y : Salary, $\quad x$: Sex, $\quad \mathbf{z}$: Rank, Degree, Year, ...

实际上，多重回归可以看作两步分析：

- 首先，从x,y中消除$\mathbf{z}$的影响：

$$x^{\perp}=x-\Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z},\quad y^{\perp}=y-\Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z}$$

（它们的相关系数即为 偏相关系数）.

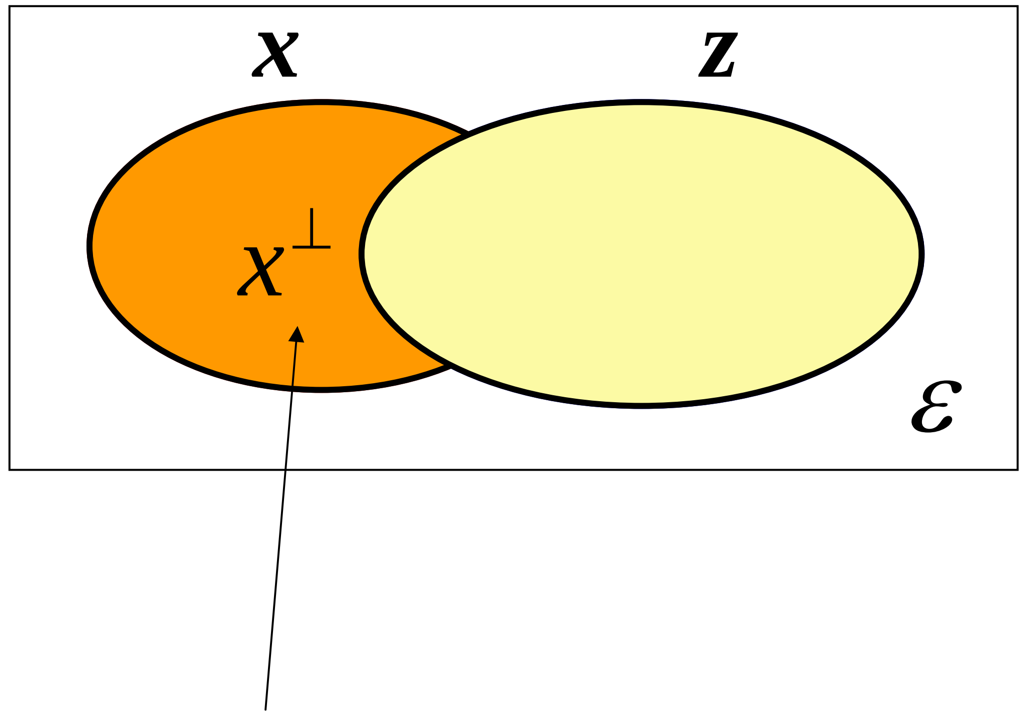
- 其次，对$y^{\perp}$(或y) 和 $x^{\perp}$假设回归模型：

$$y^{\perp}=a+bx^{\perp}+\varepsilon,\quad \varepsilon\text{与}x,\mathbf{z}\text{独立}$$

$$\Leftrightarrow\quad y-\Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z}=a+b(x-\Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}\mathbf{z})+\varepsilon$$

$$\Leftrightarrow\quad y=a+bx+c'\mathbf{z}+\varepsilon,\quad \text{其中}c'=(\Sigma_{y\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1}-b\Sigma_{x\mathbf{z}}\Sigma_{\mathbf{zz}}^{-1})$$

$y =$



控制z (给定z)时，x对y的贡献