

Mathematical Experiments

回归分析

——线性回归引例



重庆大学数学与统计学院



在科学研究和工程实践中，常常会遇到“黑箱问题”，即系统内部机理不明确，需要建立数学模型，通过观测数据，构建系统内部规律，达到预测和控制系统输出的目标。

例如

- 1 在计量经济学 (econometrics) 中，需要判断政府征收“酒类消费税”是否有助于减少交通事故次数。
- 2 在心理学 (psychology) 中，需要寻找决定学习效果的各种因素。
- 3 在数学中，需要建立因变量 y 和自变量 x 之间的函数关系 $y=f(x)$ ，即确定函数 $f(x)$ 的形式。

- 方程 $y=f(x)$ ，当 $f(x)$ 是确定性函数时，使用高等数学方法完成模型的构建、结果的预测和控制；
- 方程 $y=f(x)$ ，当 $f(x)$ 是不确定性函数时，使用概率统计方法完成函数形式的构建、结果的预测和控制。

当 $f(x)$ 是不确定性的线性函数时，称为线性回归模型。

- y 表示酒后驾车引发交通事故次数；
 $x=0$ 表示不征收酒类消费税；
 $x=1$ 表示征收酒类消费税；
 ε 表示其他因素对交通事故次数的贡献，如观测过程中的随机噪声。

- $$y = \beta_0 + \beta_1 x + \varepsilon$$

其他变量如此之多，所以我们认为扰动项是不可观测的。

- y 表示学生考试平均成绩；
 X_1 表示班级学生数；
 X_2 表示学校教师数；
 X_3 表示每日课堂教学时长；
 X_4 表示学区房价；
 X_5 表示政法教育拨款数量；
 ε 表示其他因素对考试成绩的贡献，称为扰动项。
- $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

○ $f(x)$ 含三角函数、指数函数、其他超越函数时，就是非线性回归问题。

○
$$\log(y) = \beta_0 + \beta_1 e^{\sin(x_1) + \cos(x_2)} + \beta_2 \log(x_2) + \varepsilon$$

Thanks



重庆大学数学与统计学院