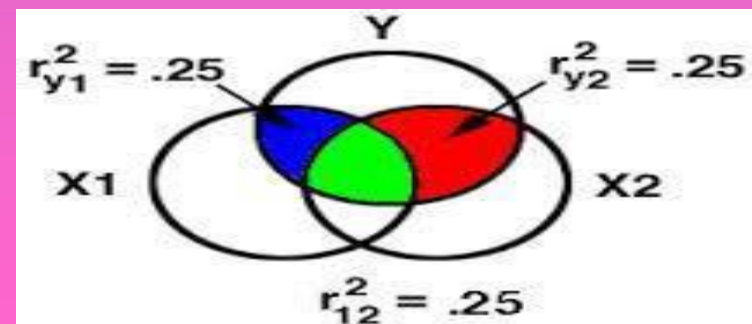


多元线性回归

- 多元线性回归是简单线性回归的直接推广，其包含一个因变量和二个或二个以上的自变量。
- 多元线性回归是研究一个因变量（Y）和多个自变量（ X_i ）之间数量上相互依存的线性关系。
- 简单线性回归的大部分内容可用于多元回归，因其基本概念是一样的。



多元线性回归方程

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + e$

参数的最小二乘估计

- 与简单回归类似，我们寻求参数 B_0 、 B_1 、 B_2 和 B_p 的适宜估计数值 b_0 、 b_1 、 b_2 和 b_p ，使实际观察值和回归方程估计值之间残差平方和最小，

$$\begin{aligned} \text{即 } Q &= \sum (y_i - \hat{y}_i)^2 \\ &= \sum (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \cdots - b_p x_{pi})^2 \end{aligned}$$

对 b_0 、 b_1 、 \dots 、 b_p 分别求偏导数，令偏导数为零可获得 $P+1$ 个正规方程，求解正规方程可得待估参数值。

多元线性回归方程的作用

- 因素分析
- 调整混杂因素的作用
- 统计预测



回归分析中的数据要求

- **因变量是连续性的数值**
- n 足够大，至少应是自变量个数的5倍；
- 自变量 x 与应变量 y 之间存在线性关系；
- 正态性：多元正态分布。
- 等方差：对于所有的自变量 x ，残差 e 的条件方差为 σ^2 ，且 σ 为常数；
- 独立性：在给定自变量 x 的条件下，残差 e 的条件期望值为0（本假设又称零均值假设）；
- 无奇异点（outliers）；
- 无共线性：自变量 x 之间相互独立。

分类变量的输入转换

- 虚设代码(dummy coding)

- 虚设代码数=变量的组数-1

例子

- 教育程度-小学，中学，大学

教育程度——小学； 教育程度——中学

小学	1	0
中学	0	1
大学	0	0

- 虚设代码数=3-1=2； 教育程度——小学； 教育程度——
中学

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/097032016032006032>