

中级应用统计学

相关分析和回归分析

张晨峰

华东理工大学商学院

2016年9月19日

2 相关分析和回归分析

主要内容

- 线性相关分析
- 简单回归分析
- 多元回归分析

2.1 线性相关分析

一些例子

- 高脂肪的食物比低脂肪的食物含有更多的热量吗？
- 父母的身高与他们孩子的身高有关吗？

相关分析

相关分析描述的是两个变量间关系的方向和强度。

2.1 线性相关分析

散点图

对于两个数值变量，我们总是可以用散点图来开始分析它们之间的关系。

2.1 线性相关分析

相关强度的度量

- Pearson相关系数，又叫相关系数或线性相关系数
- Kendall τ 相关系数
- Spearman秩相关系数

2.1 线性相关分析

定义 (总体相关系数)

两个变量间的皮尔逊相关系数定义为两个变量间的协方差和标准差的商

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

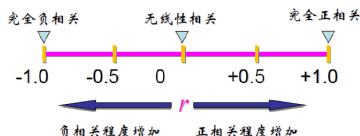
定义 (样本相关系数)

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

2.1 线性相关分析

相关系数

- 相关系数的值域范围是 $[-1, 1]$
- 相关系数的绝对值反映相关的强度
- 相关系数的正负号反映相关的方向



2.1 线性相关分析

相关系数的显著性检验

- 原假设: $\rho = 0$
- t检验: $t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$, $df = N - 2$
- 需要注意的是, 如果要检验其为某一具体值, 则需要转换

2.1 相关分析

Network Diversity and Economic Development

