解决多重共线性的代码实现

检测方法以及解决方法

小胖

目录

ONE 两个变量的检测方法

one-way ANOVA

TWO 多变量的检测方法

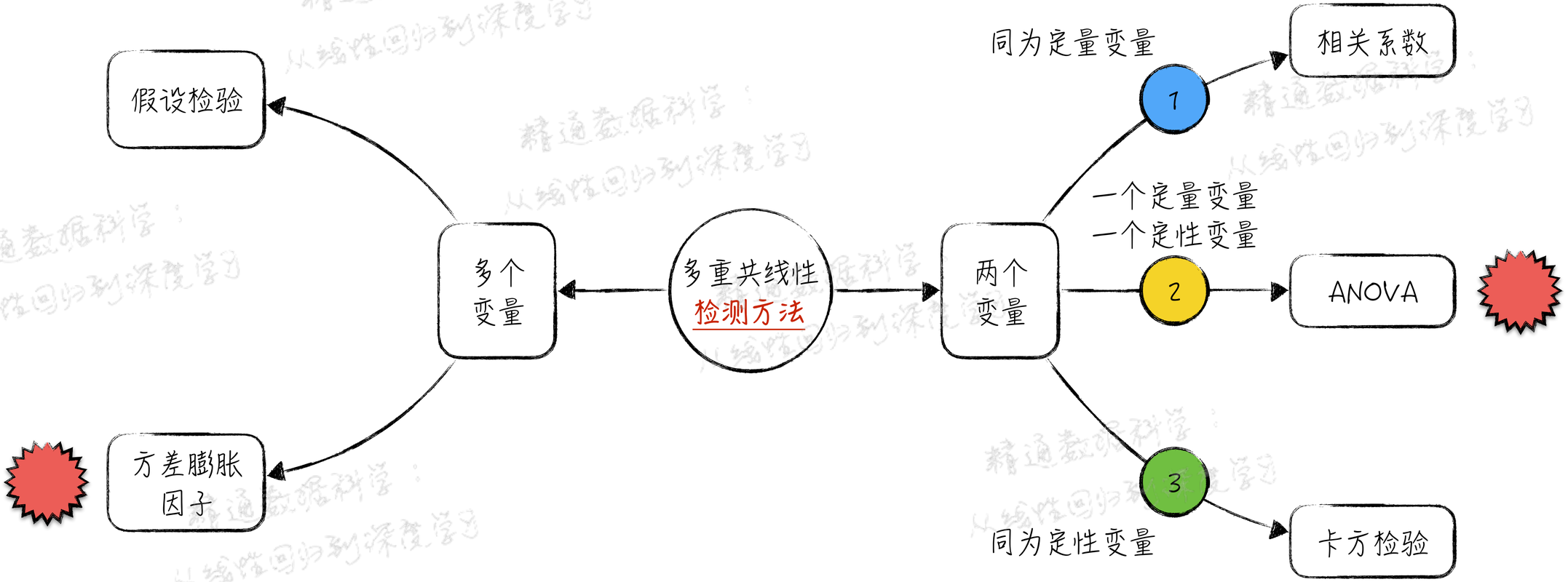
方差膨胀因子

THREE 虚拟变量陷阱

一个简单的例子

两个变量的检测方法

整体回顾

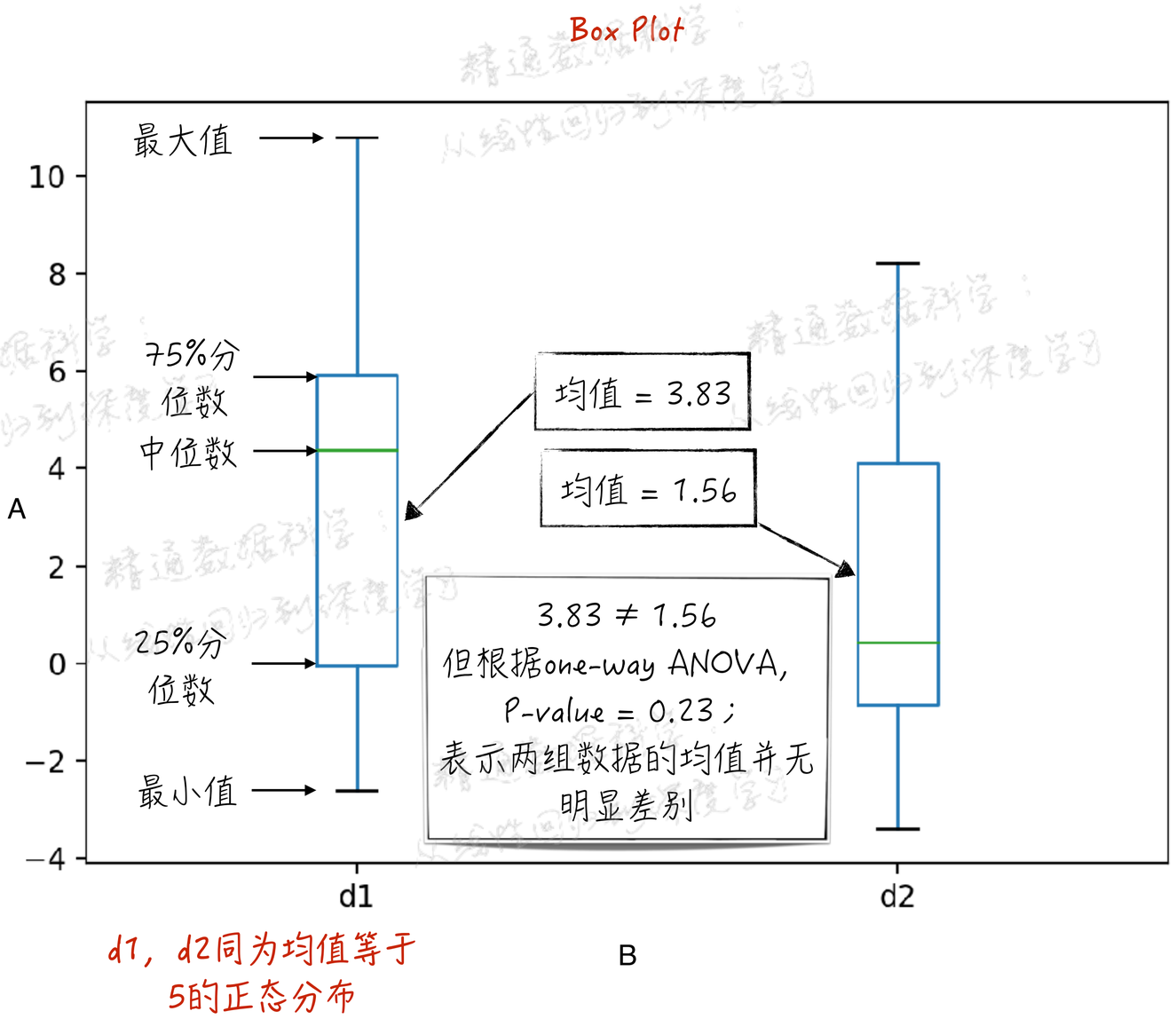


两个变量的检测方法

one-way ANOVA

one-way ANOVA：多组数据的均值是否**显著**不同

ANOVA：analysis of variance



两个变量的检测方法

one-way ANOVA

ANOVA：analysis of variance

多重共线性的检测

假设A是定量变量、B为定性变量（共有k个分类）

变量A按B的取值分为k组

定义技术指标 η^2

变量A的全部方差 S_{total}

变量A的组内方差 S_{within}

变量A的组间方差 $S_{between}$

$$\eta^2 = \frac{S_{between}}{S_{total}}$$

η^2 越接近1，共线性越严重

目录

ONE 两个变量的检测方法

one-way ANOVA

TWO 多变量的检测方法

方差膨胀因子

THREE 虚拟变量陷阱

一个简单的例子

多变量的检测方法

方差膨胀因子

假设如下的线性回归模型： $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$

随机扰动项的方差

数学上可以证明： $\hat{Var}(\hat{\beta}_1) = \frac{\sigma^2}{Var(x_1)} \frac{1}{1 - R_1^2}$

下面线性回归模型的决定系数

x1能多大程度被其他变量解释： $x_1 = \alpha_0 + \alpha_2 x_2 + \ldots + \alpha_k x_k + e$

计算VIF时，不要忘记在数据里加入常数项

定义方差膨胀因子： $VIF_1 = \frac{1}{1 - R_1^2}$

只与数据相关，与使用的模型无关

目录

ONE 两个变量的检测方法

one-way ANOVA

TWO 多变量的检测方法

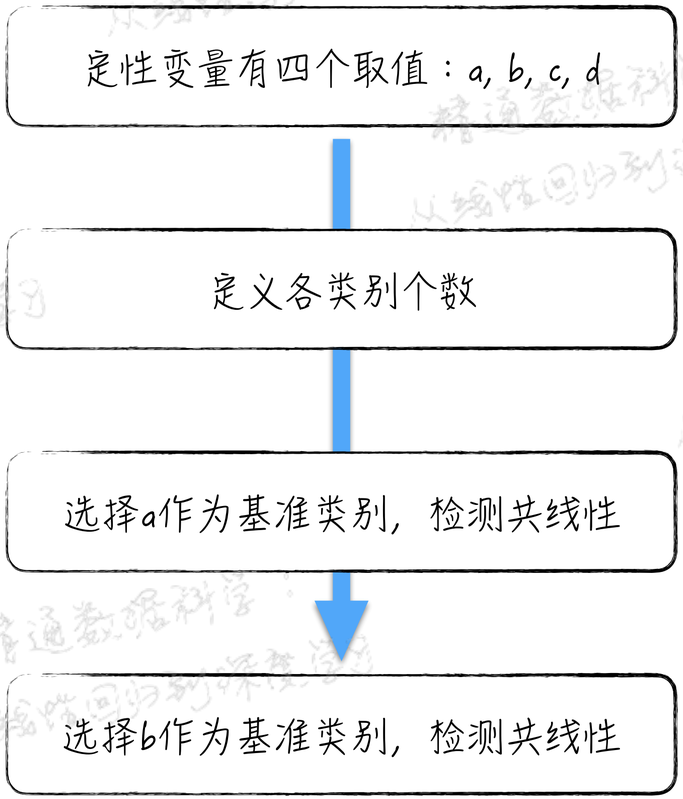
方差膨胀因子

THREE 虚拟变量陷阱

一个简单的例子

虚拟变量陷阱

这里是副标题文字



类别	个数
a	5
b	100
c	70
d	20

THANK YOU