

# 数据挖掘 第2周课后作业: 基本统计描述

---

## 2.1 假设所分析的数据包括属性age，它在数据元组中的值为：

13, 16, 52, 16, 20, 20, 21, 22, 25, 35, 25, 25, 30, 33, 33, 19, 35, 22, 35, 35, 15, 35, 36, 40, 45, 25, 46, 70

- (1) 该数据的均值是多少？中位数是什么？
- (2) 该数据的众数是什么？讨论数据的模态。
- (3) 该数据的中列数是多少？
- (4) 你能找出该数据的第一个四分位数和第三个四分位数吗？
- (5) 给出该数据的五数概括。
- (6) 绘制该数据的盒图。
- (7) 分位数-分位数图与分位数图有什么不同？

(1) 均值为30.5，中位数是按顺序排列的一组数据中居于中间位置的数，对于给定的这组数据，中位数为 $(30+32)/2=31$ ；

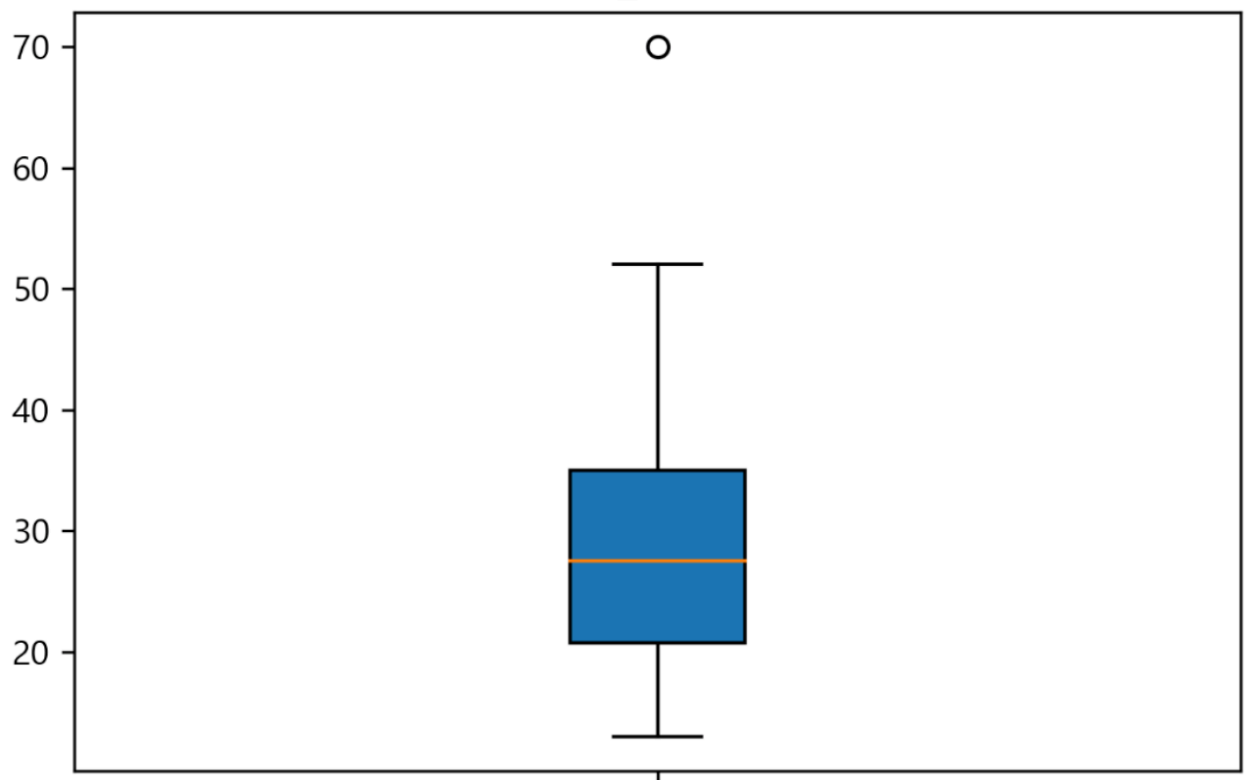
(2) 该数据的众数是出现次数最多的数据，为35（出现了5次），为单模态，右倾斜；

(3) 中列数为28.5；

(4) 第一个四分位数为20.75，第三个四分位数为35；

(5) 五数概括：13, 20.75, 27.5, 35, 70；

(6) 盒图如下：



(7) 分位数-分位数图是概率图，而分位数图用于观察单变量，使得用户可以观察从一个分布到另一个分布是否有漂移。

## 2.2 假设医院对18个随机挑选的成年人检查年龄和身体肥胖，得到如下结果：

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2

age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

(1) 计算age和%fat的均值、中位数和标准差。

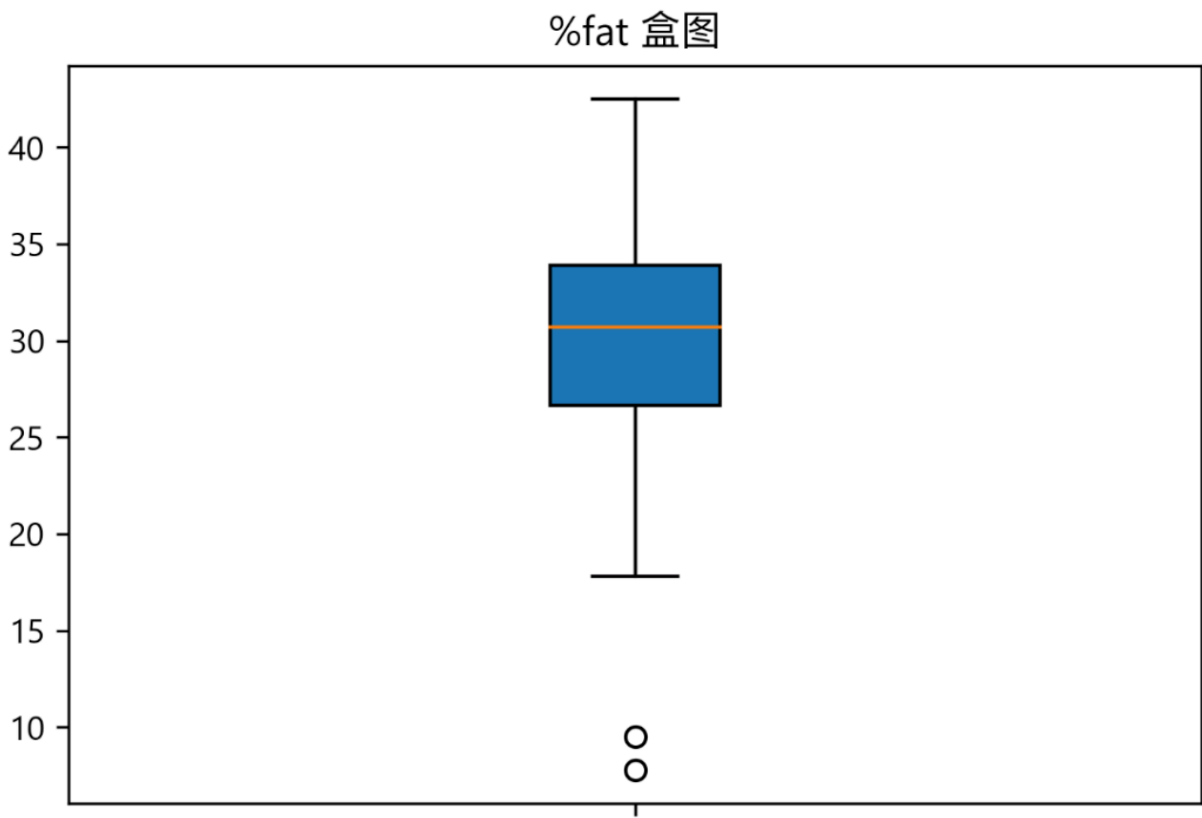
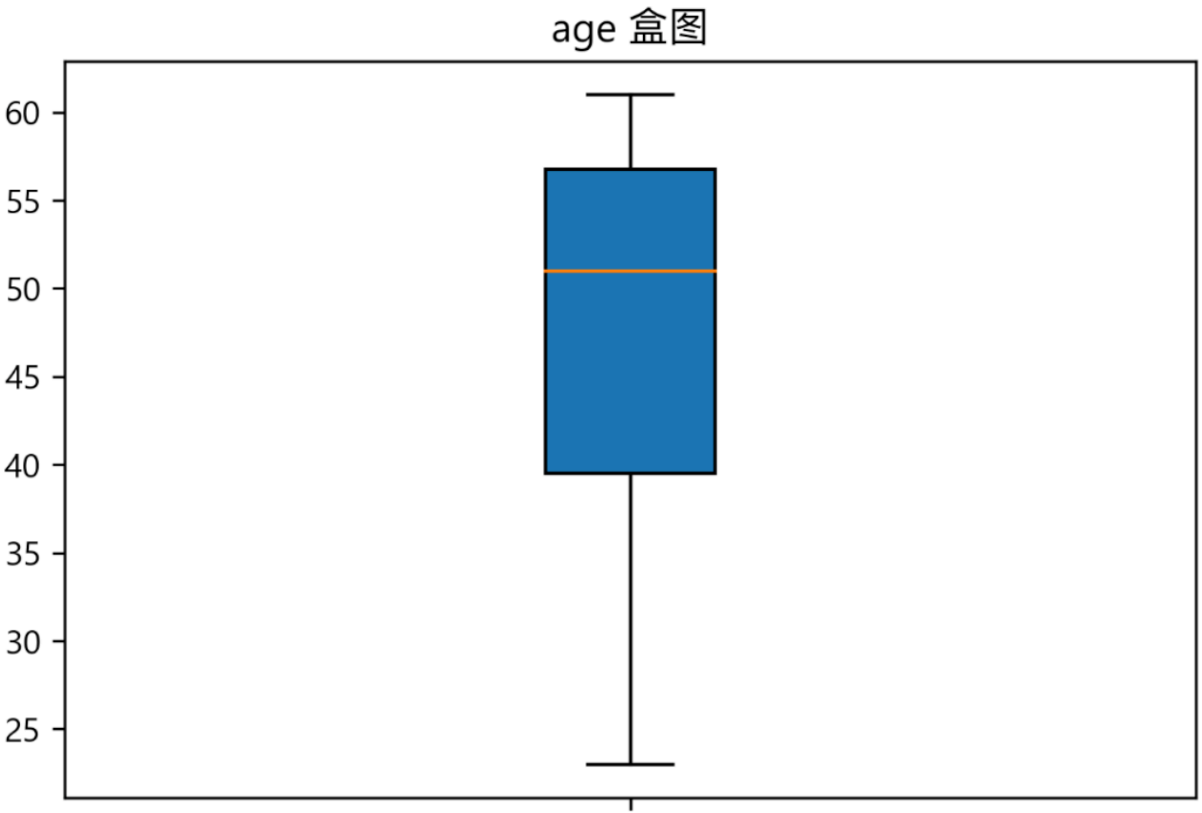
(2) 绘制age和%fat的盒图。

(3) 绘制基于这两个变量的散点图和q-q图。

(1)

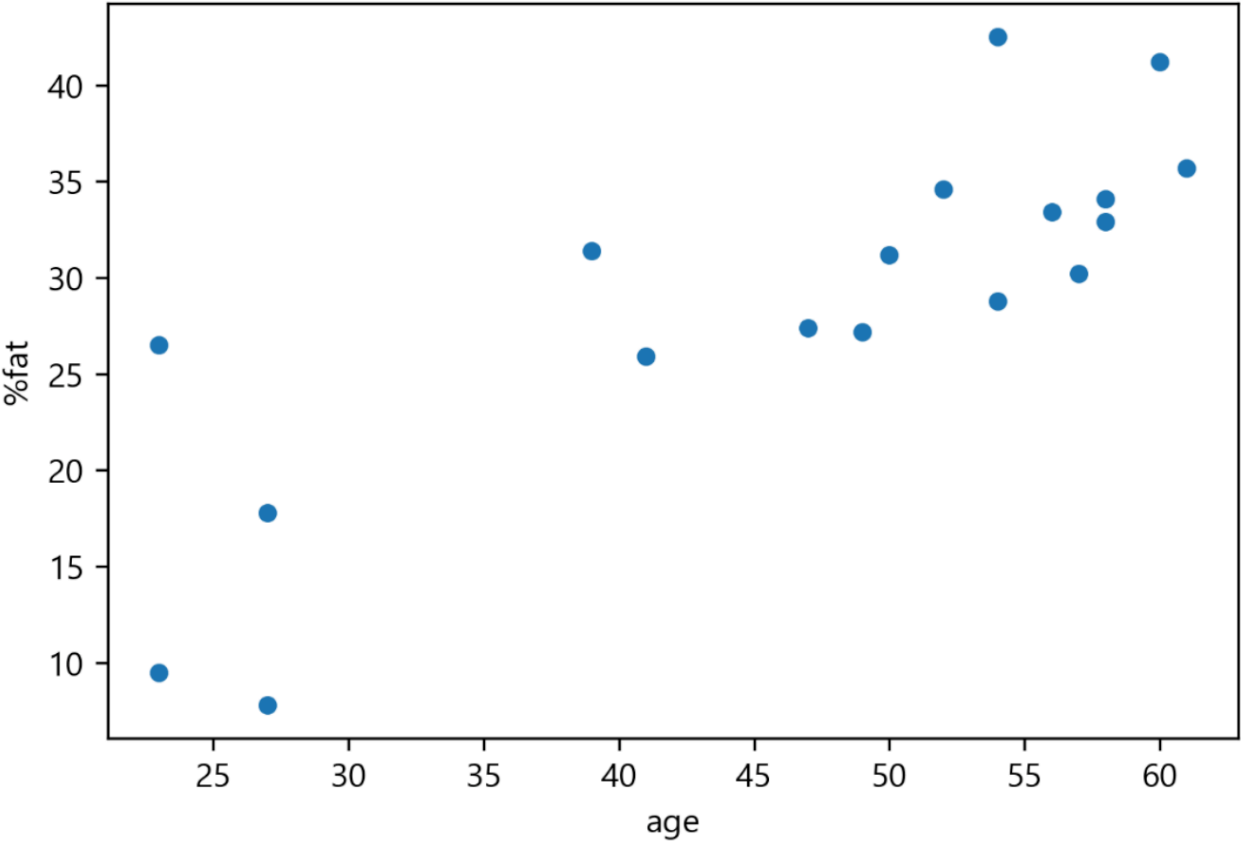
项目	均值	中位数	标准差
age	46.444	51	13.219
%fat	28.783	30.7	9.254

(2)



(3)

散点图



Q-Q 图

