

5.1 简介

几乎任何一个数据集都含缺失数据。缺失数据有多种表示形式。在数据库中，缺失数据表示为 NULL 值；在某些编程语言中使用 NA 表示。根据数据的来源，缺失值可能是空字符串（''）或数值（比如 88 或 99）。在 Pandas 中使用 NaN 表示缺失值。

概念一览

预备知识：

- a. 导入库；
- b. 数据分片和索引；
- c. 使用函数和方法；
- d. 使用函数参数。

目标

- (1) 认识缺失值；
- (2) 创建缺失值；
- (3) 对缺失值重新赋值和计算。

5.2 何为 NaN 值

Pandas 中的 NaN 值来自 NumPy 库。在 Python 中，缺失值有几种表示形式：NaN、NAN 或 nan，但它们都是等同的。附录 H 介绍了导入这些缺失值的方法。

```
# 从 NumPy 库导入缺失值
from numpy import NaN, NAN, nan
```

缺失值和其他类型的数据不同，实际上它们无甚意义。数据缺失了，也就无所谓等于或不等于。NaN 也不等同于 0 或者空字符串（''）。