

統計

[数学][R]

データ解析に重要な 4 つの S

項目	内容
採集	入力、転記ミスはないか
視覚化	数値だけで判断しない。グラフ、ヒストグラム、パレート図、散布図など
層別	様々な角度から見る。男女別、年代別、月別、曜日別など
相関	2 種類のデータの関係性を見つける。年齢と年収、経済成長率と失業率など

Python

- 基本統計量、ヒストグラム
- 基本統計量、分散と標準偏差
- 正規分布、確率密度関数
- 散布図と相関分析 (1)
- 散布図と相関分析 (2) 単回帰分析
- 重回帰分析

Tips

- ・ ギリシャ文字
- ・ ヒストグラム範囲の決め方

クーポン収集問題

- ・ 6 面サイコロの全ての目が出るまでにサイコロを平均で何回振る必要があるか

```
>>> p = 6
>>> sum([ 1.0 / (x/float(p)) for x in range(1,p+1)])
14.7
```