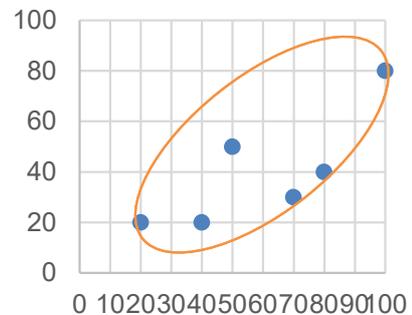


## 内容

### 1. データサイエンスとは

- ・データとは
- ・分布の特定の位置を示す統計量
- ・代表値（平均値、中央値、最頻値）
- ・分布のばらつきを示す統計量
- ・散布度（分散、標準偏差、四分位範囲）
- ・データの可視化（グラフ）
- ・ヒストグラム、箱ひげ図、散布図、棒グラフ、円グラフ
- ・オープンデータ（e-Stat）

散布図



### 2. データサイエンスの手法

- ・統計学の歴史
- ・機械学習と多変量解析
- ・教師あり学習 回帰  
回帰と相関、単回帰、重回帰
- ・教師あり学習 分類  
k近傍法、決定木、ランダムフォレスト、勾配ブースティング木
- ・教師なし学習 次元削減  
主成分分析
- ・教師なし学習 クラスタリング  
階層クラスター分析、非階層クラスター分析
- ・ニューラルネットワーク

### 3. ベイズ推定

- ・確率の公理、条件付き確率、全確率の定理、ベイズの定理

### 4. Kaggle タイタニック生存予測問題（分類）

- ・分析の流れ、注意点

### 5. 演習（Excel, Python）

- （Excel）
- ・ピボットテーブル
- ・マンションの家賃予測（回帰）
- （Python）
- ・得点データの主成分分析（次元削減）
- ・クラスター分析（クラスタリング）