

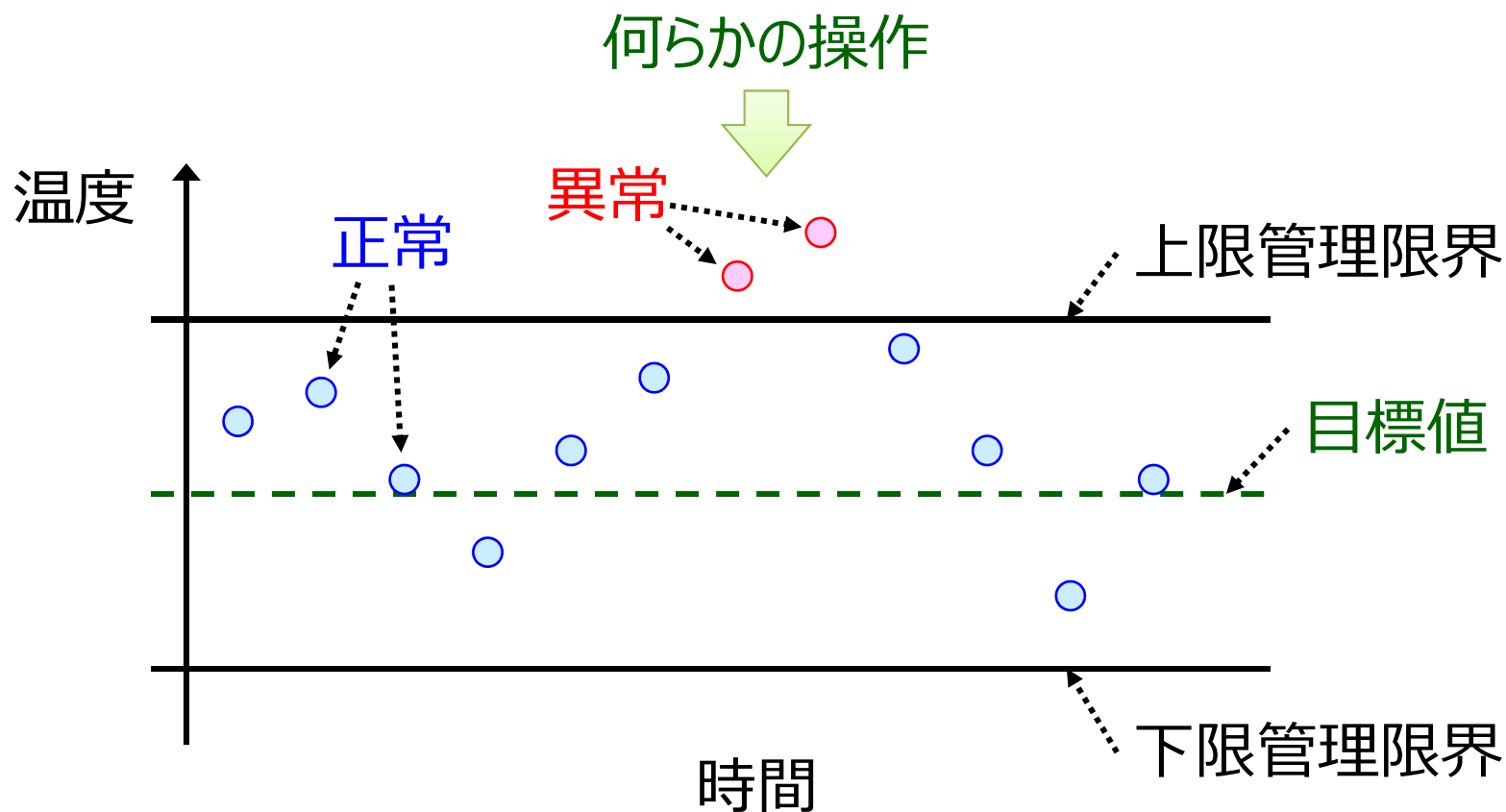
T²統計量・Q統計量

明治大学 理工学部 応用化学科
データ化学工学研究室 金子 弘昌

T²統計量・Q統計量とは？

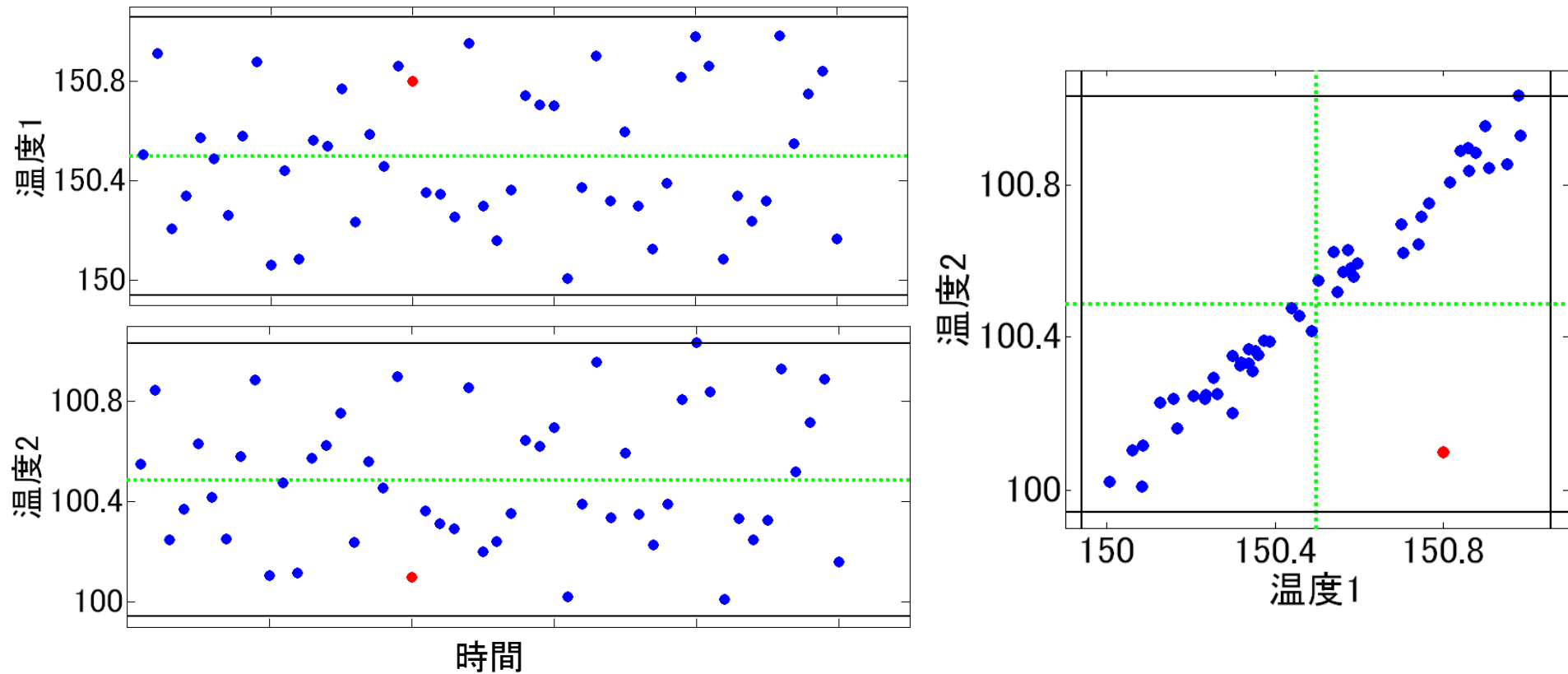
- ✓データに基づいた異常検出・異常診断に用いられる
- ✓モデルの適用範囲 (Applicability Domain, AD) の設定にも利用できる
 - <https://datachemeng.com/applicabilitydomain/>
- ✓T²統計量・Q統計量は主成分分析 (Principal Component Analysis, PCA) に基づく
 - <https://datachemeng.com/principalcomponentanalysis/>
- ✓T²統計量は、標準化 (オートスケーリング) した第 m 主成分までの、原点からの距離の二乗
- ✓Q統計量は、第 m 主成分までで表せなかった部分の大きさ

プロセスの異常検出・制御



シューハート管理図

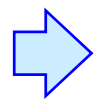
2つのプロセス変数の相関が強い場合



赤点は分布から外れているにもかかわらず異常値として検出されない

多変量プロセス管理

- ✓プラントでは非常に多くのプロセス変数が測定
 - 一つのプロセス変数の誤アラーム率が **0.01%** でも100変数のどれかに誤アラームがある確率は **約10%** !
 - 4変数以上 (4次元以上) のデータの様子は確認できない



多変量プロセス管理

(Multivariate Statistical Process Control, MSPC)

– 主成分分析

– 独立成分分析

など

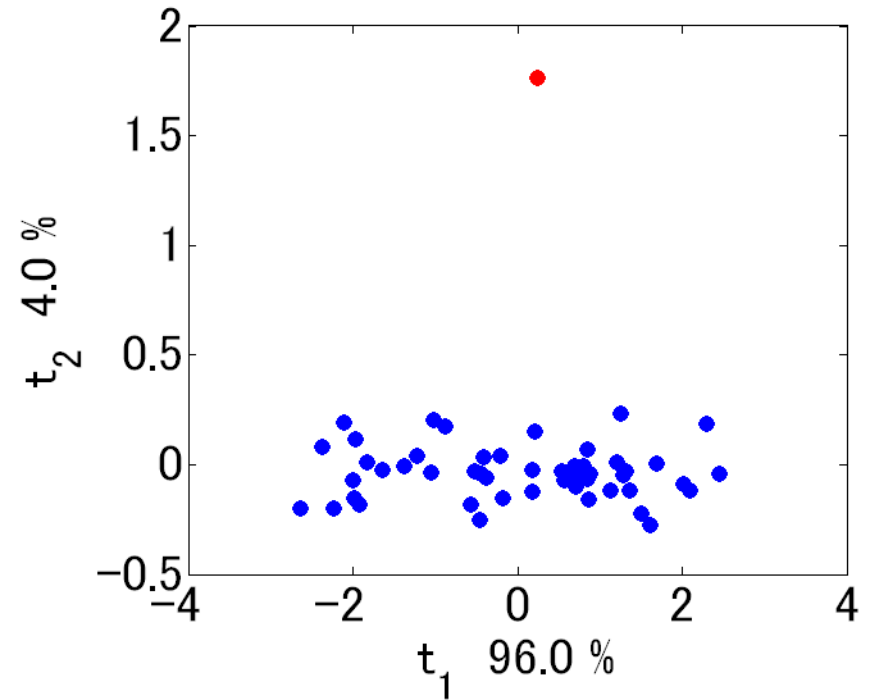
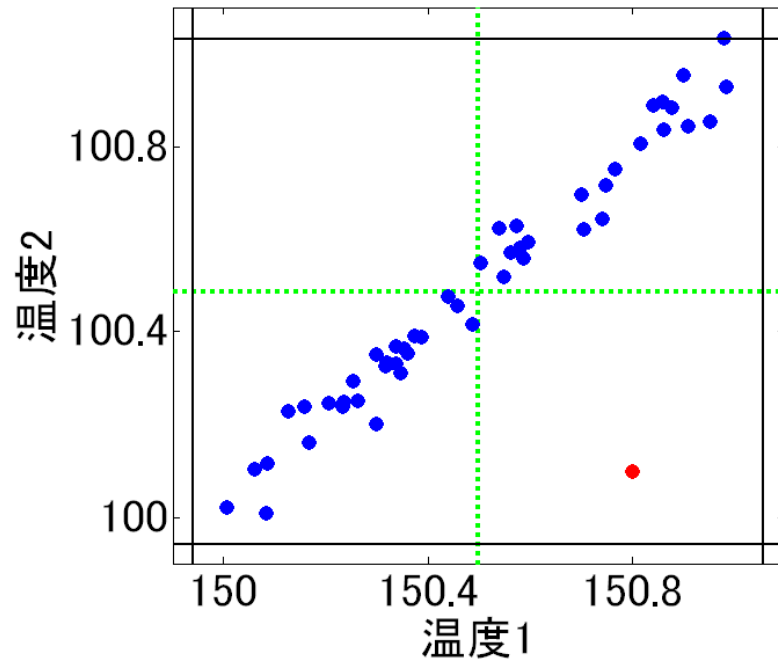
主成分分析(PCA)

✓主成分分析 (Principal Component Analysis, PCA)

- 見える化 (可視化) する手法
- 多変量 (多次元) のデータセットを低次元化する方法
- データセットのもつ情報量をなるべく失わないように元の次元からより低い次元でデータセットを表現
 - “より低い次元” を 2 次元にすれば可視化を達成
- 軸を回転 (+ 反転) させる

✓詳しくは、<https://atachemeng.com/principalcomponentanalysis/>

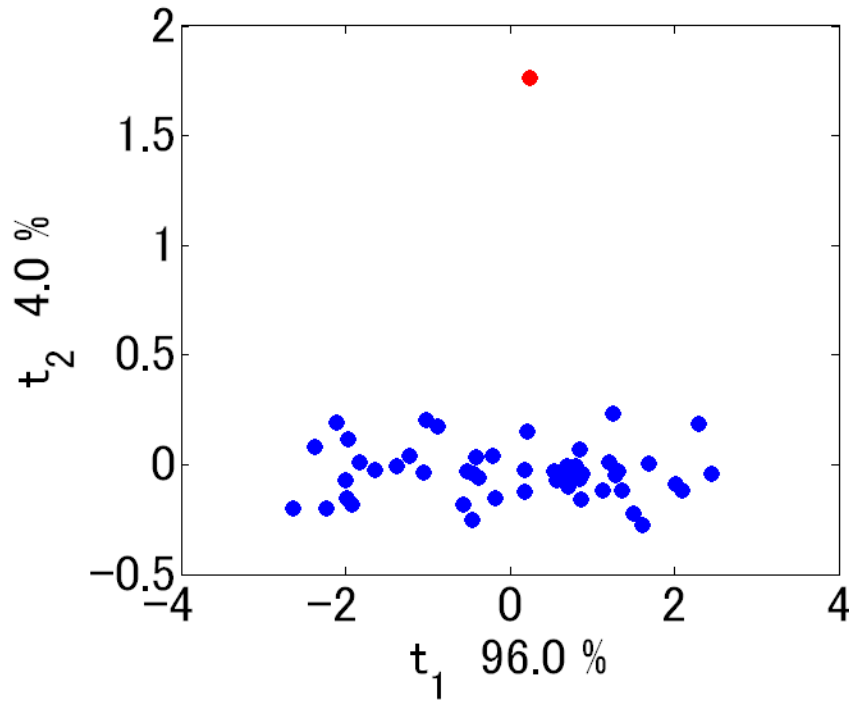
PCA の例



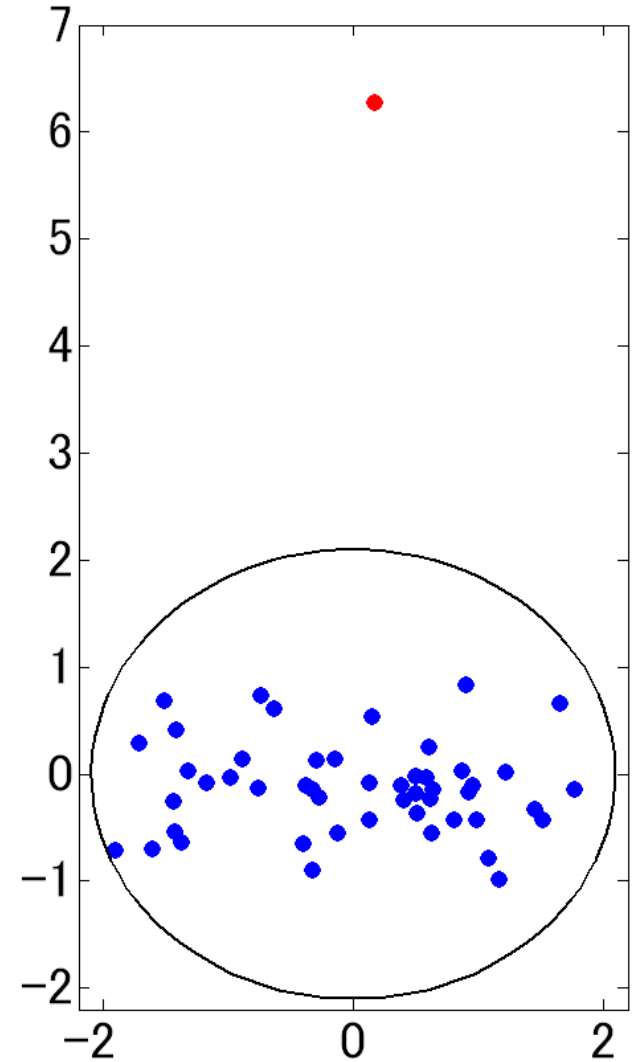
第2主成分軸 (t_2) を見れば、異常値を検出できそう！

主成分軸がもっと多いときはどうする？

T²統計量



標準化
(各主成分を
標準偏差で
割る)



T²統計量の計算

あるサンプルの T²統計量

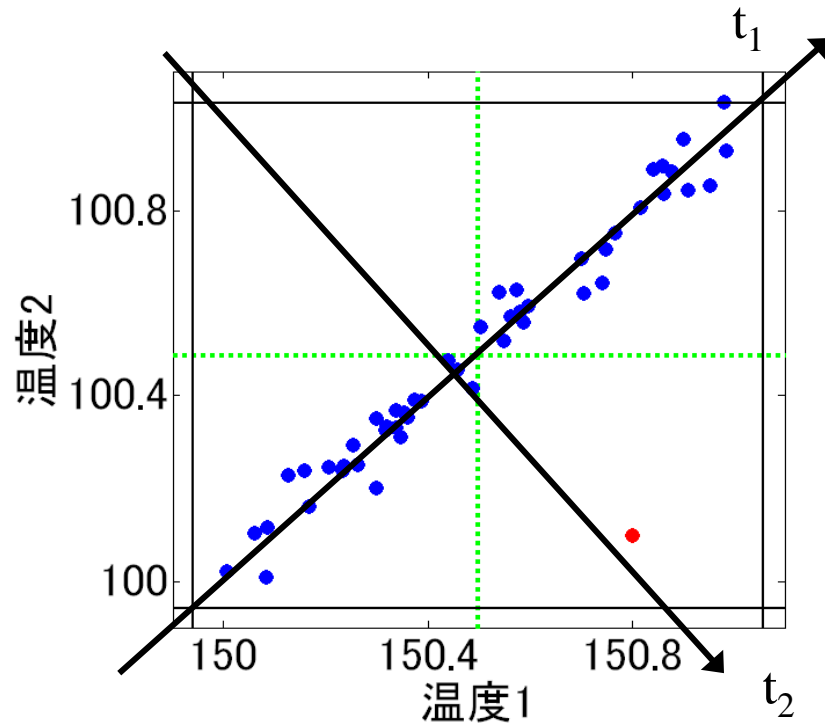
$$T^2 = \sum_{i=1}^m \left(\frac{t_i}{\sigma_i} \right)^2$$

t_i : 第 i 主成分のスコア

σ_i : 第 i 主成分の標準偏差

m : 考慮する主成分の数

Q統計量：PCAで表現できない部分



1 成分まで使用する場合、

Q統計量 = t_1 軸との距離の二乗

2 成分まで使用する場合、

Q統計量 = t_1 軸, t_2 軸の平面からの
距離の二乗

⋮

あるサンプルの Q統計量

$$Q = \sum_{i=1}^n (x_i - x_{\text{EST},i})^2$$

n : 変数の数

x_i : i 番目の変数の値

$x_{\text{EST},i}$: PCAによる i 番目の変数の推定値
(第 m 主成分までで逆写像されたサンプル点の
 i 番目の変数に対応する値)

逆写像の詳細については

<https://atachemeng.com/principalcomponentanalysis/>

T²統計量・Q統計量の使い方

- ✓ T²統計量・Q統計量のどちらかが閾値を越えたときに**異常値**とする
- ✓ 閾値をどう決めるか？
 - 3σ法において正規分布に従うとき±3σ以内に入る確率は99.7%である
 - PCAモデルを構築したときのデータセット (すべて正常サンプル) を用いて、T²統計量 (Q統計量) の小さい順にサンプルを並び替え、99.7% に対応するサンプルのT²統計量(Q統計量) の値を閾値とする
 - たとえば 1000サンプルあったときは、997番目のサンプルのT²統計量(Q統計量) の値

モデルの適用範囲 (AD)

✓ T^2 統計量・Q統計量のどちらかが閾値を越えたときに

モデルの適用範囲外とする

✓ モデルの適用範囲についてはこちら

<https://datachemeng.com/applicabilitydomain/>

Q統計量で異常に関与する変数の診断

13

変数ごとに x_i と $x_{\text{EST},i}$ との差をみて、
それが大きい変数が異常に関与している、とする