

情報処理応用B 第8回

藤田 一寿

■ 出席について（確認）

- 出席は課題提出で確認しています.
- 講義に出席しているから課題を受け取れる.
- 課題を提出する.
- つまり、課題提出が可能なのは講義に出席しているから.
- よって、出席している講義に配布した課題だけが提出されていると判断しています.

問題解決と統計処理

p194

■ 分析と改善

- 問題が起こったときにどう解決すればよいのか.
- 現状を分析し改善する.
- 現状分析の手法
 - 統計処理
- 改善の手法
 - PDCAサイクル

■ PDCAサイクル

- 改善のための4つの手順(PDCAサイクル)

- ① 計画(P: Plan)

- 目的を決め、達成に必要な計画を設定

- ② 実行(D: Do)

- 計画に基づき実施

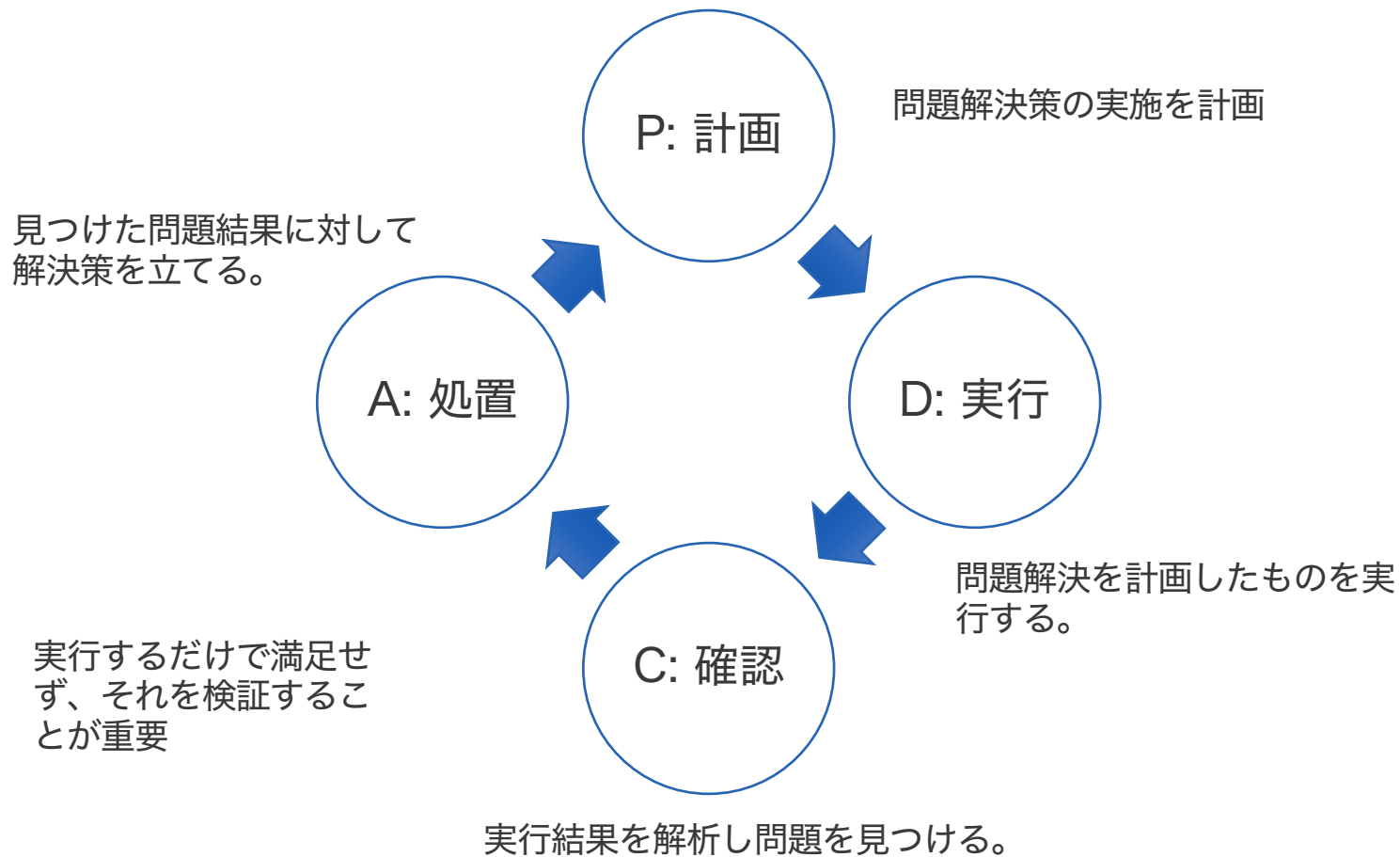
- ③ 確認(C: Check)

- 実施の結果を調べ評価

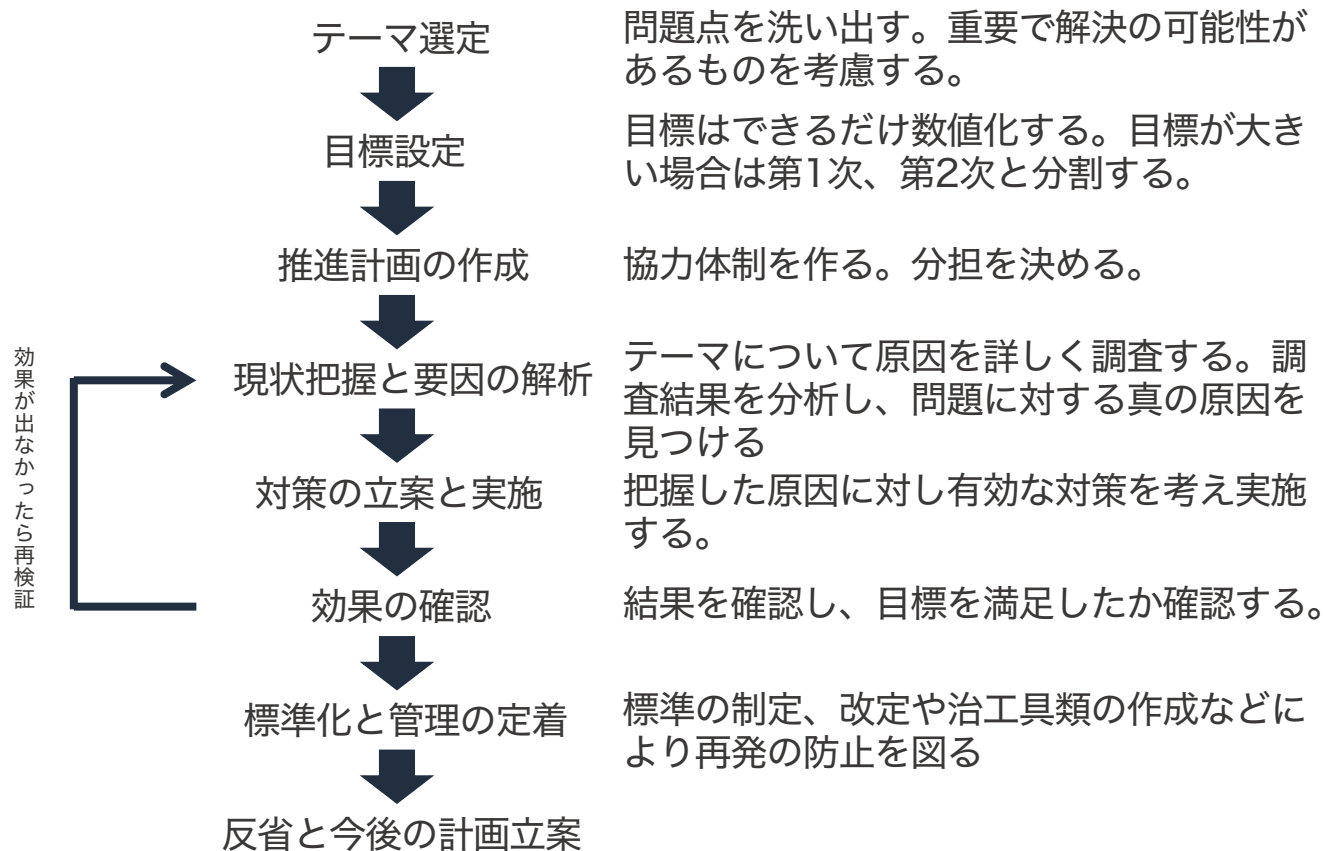
- ④ 処置(A: Action) 処置

- 必要により適切な処置

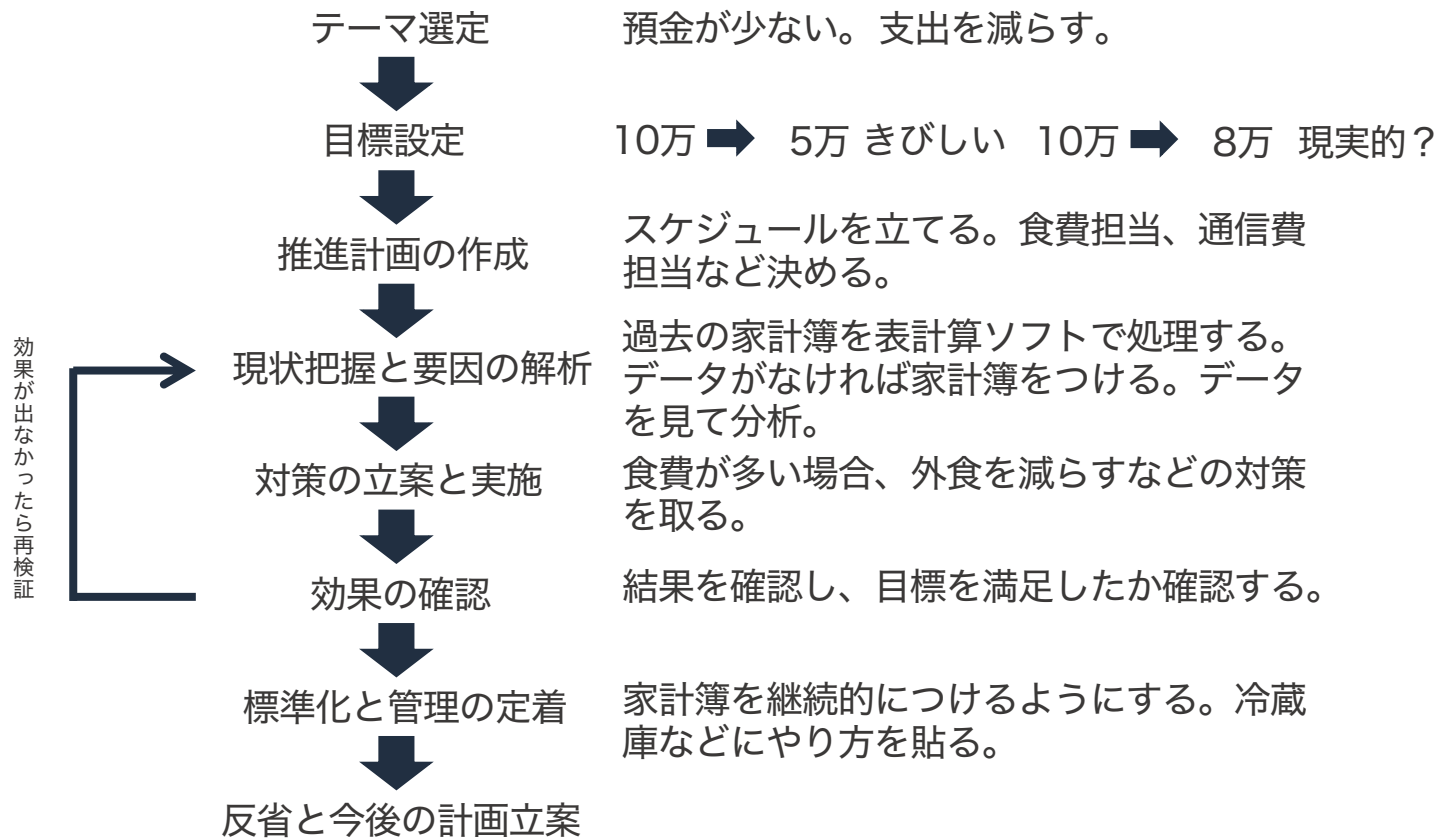
■ PDCAサイクル



■ 改善のアプローチ



■ 家計支出の例



統計の基礎

問題を解決するには、経験や感だけではなく、データや理論を用いて現状を認識する必要がある。

数値化することが重要

■ 問題解決で統計やグラフが必要なわけ

- 現状を数値化し評価したい
- すべての製品をチェックできないが製造工程全体を評価したい
- 現状を可視化してわかりやすく理解したい
- 図で表すことで問題点を発見しやすくしたい
- など

■ なぜ統計が必要か

- 無数にあるデータ一つ一つをチェックできない
- 個々のデータをみるだけでは全体の傾向がつかめない



統計的手法を用いる

■ 統計は様々な分野で使われる

- 製造業では品質管理や販売予測など
- パターン認識は確率統計の理論が用いられている

■ 統計で用いる基礎的な数値（統計量）

- データ数
- 最大、最小、メディアン(中央値)
- 平均
- 平方和
- 分散、標準偏差(ばらつき具合)

■ 最大値、最小値、メディアン

- 最大値
 - 最も大きい値
- 最小値
 - 最も小さい値
- メディアン
 - 順番的に中央の値
 - 平均より中央値の方が適切な場合もある。平均は値が大きいデータに引っ張られるため。

データが奇数個の場合



データが偶数個の場合



■ 平均

- N 個のデータ x_1, x_2, \dots, x_N の平均は下記のように表される。

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

■ 分散、標準偏差

- 分布のばらつき具合を表す指標
 - 分散

$$V = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x})^2$$

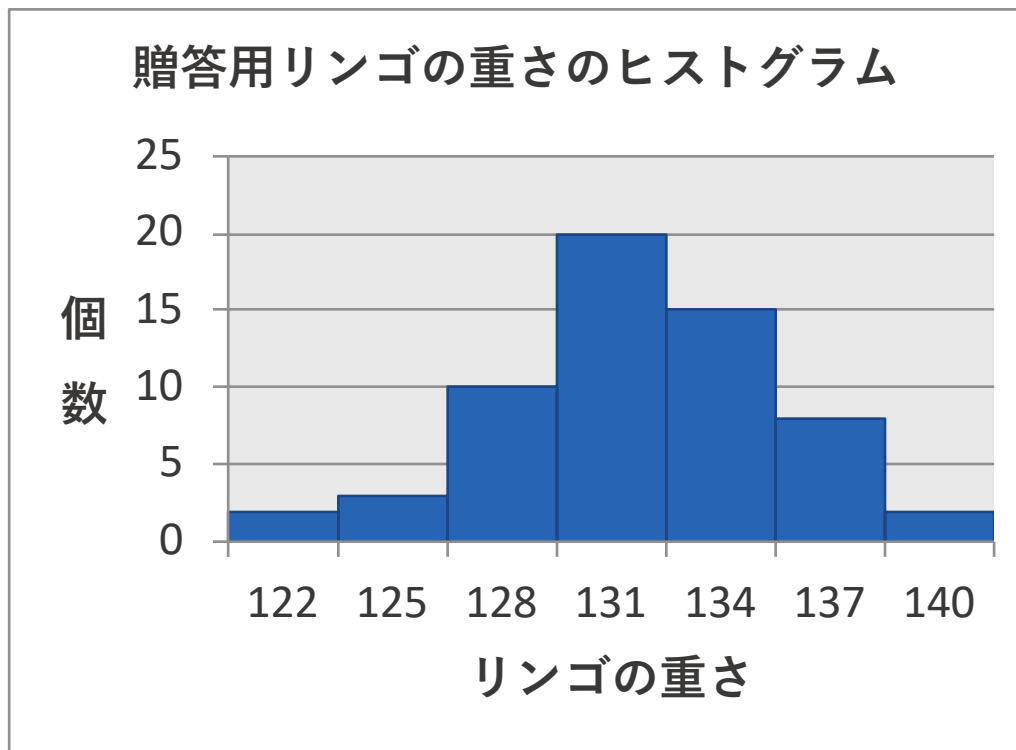
- 標準偏差

$$S = \sqrt{V}$$

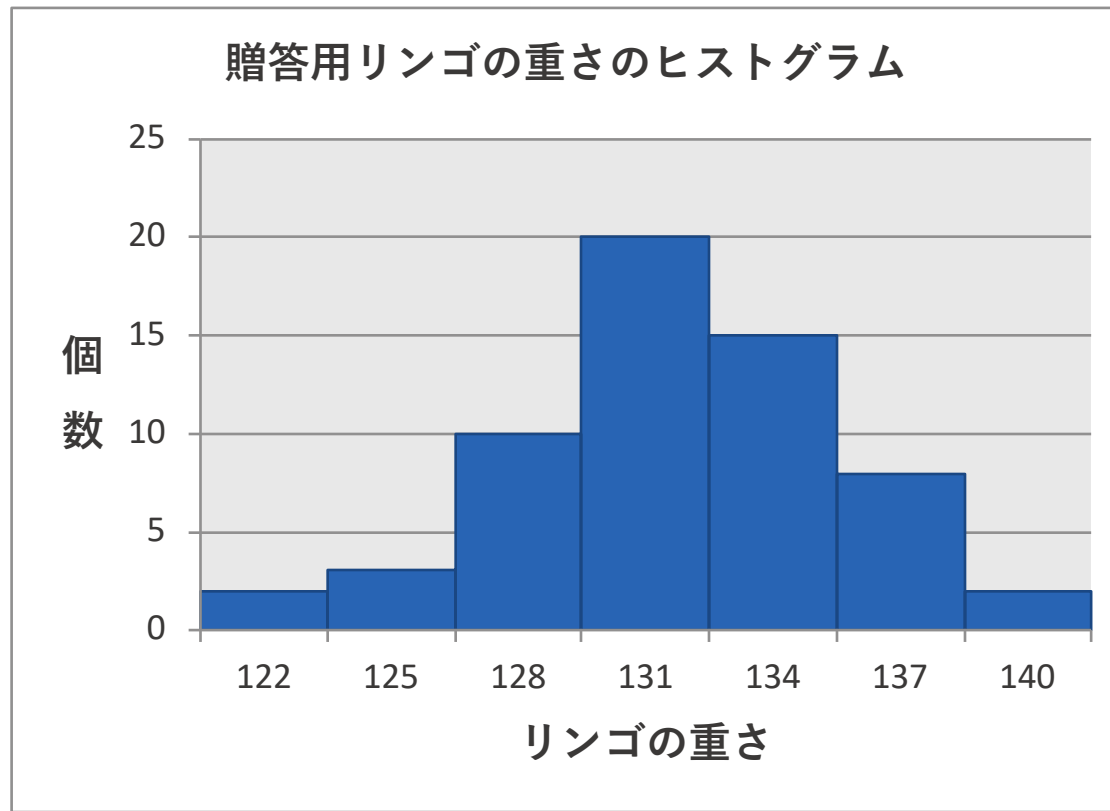
(σ)

■ ヒストグラム(度数分布)

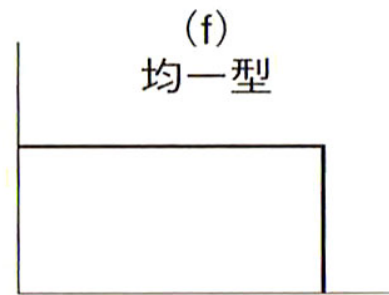
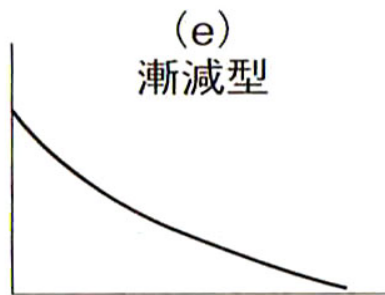
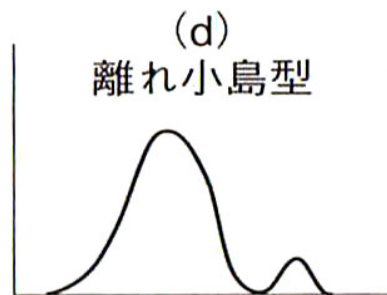
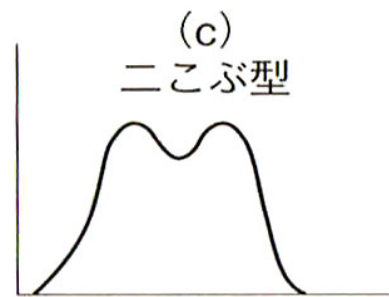
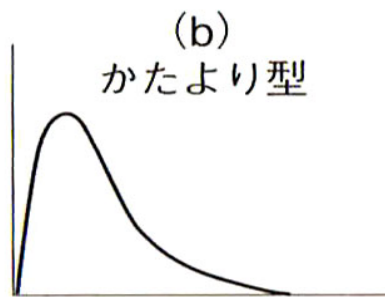
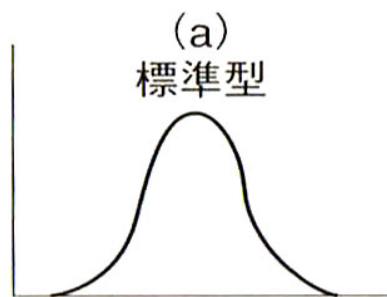
- データの中心やばらつきの大きさの把握
- 規格との比較や異常値の把握（分布の形状が規定値かどうか）



リンゴの重さのデータ表(g)					
133	130	127	121	137	130
132	130	129	130	130	137
135	133	121	129	132	130
140	133	132	129	129	132
126	132	132	127	129	129
130	124	135	137	127	132
126	129	130	135	137	132
130	130	127	133	135	124
126	127	130	132	133	126
124	127	140	130	132	129

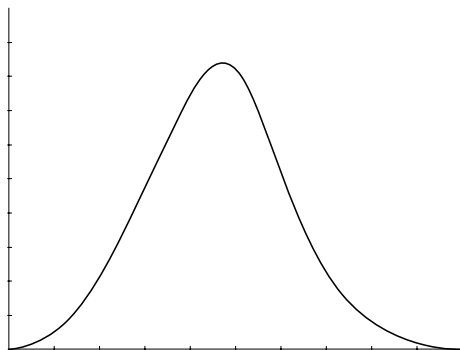


■ 分布の形状



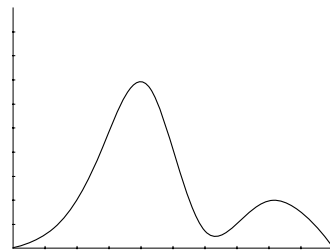
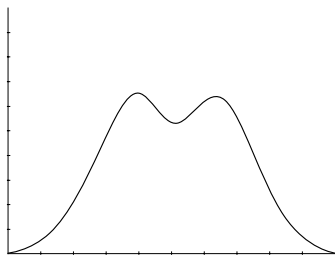
■ 正規分布

- 最も基礎的な分布の形
 - 中心極限定理によりサンプルの数を多くすると近似的に正規分布になる
- 正規分布と呼ぶ
 - 標準型、一般型、ベル型、ガウス分布などと呼ばれることもある
- 何かを測定した場合、誤差を含めこの分布になる。



■ ふたこぶ型

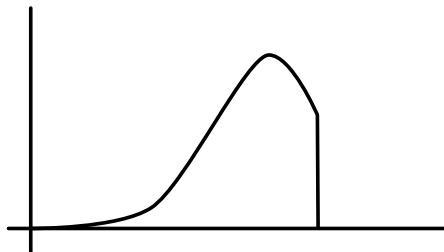
- 複数の要素を含む場合に生じる。
 - 成績の分布
 - 理解している人の集団と理解していない集団のそれぞれの分布がある場合
 - 製造装置の一部が違う規格で作っている場合
- このような分布が見られた場合、その原因を探らなければならない。



■ その他

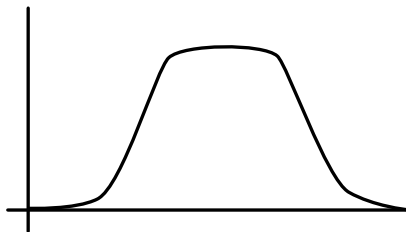
- 絶壁型

- 規格以下もしくはは以上のものを選別して取り除いたときに現れる分布。



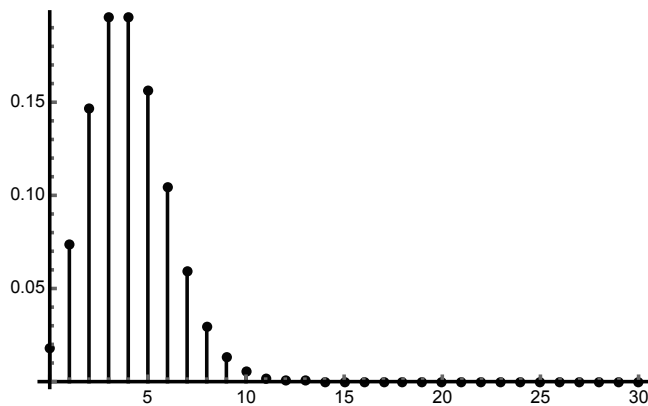
- 高原型

- ふたこぶ型の一種
- 平均値が少し異なるいくつかの分布が混在したときに現れる分布。
- 層別して原因を探る必要あり。



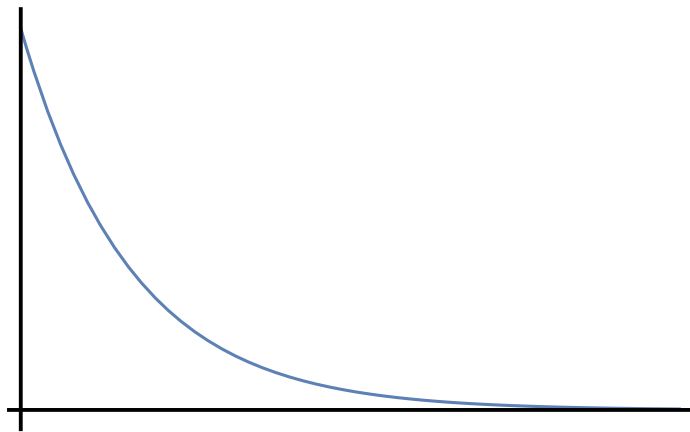
■ ポアソン分布

- 正規分布が偏ったもの。限界値(値が負にならないなど)があった場合みられる。
- ポアソン分布と呼ぶ。
 - 偏り型などと呼ばれることもある。
- 交通事故件数、大量生産の不良品件数、火災件数など



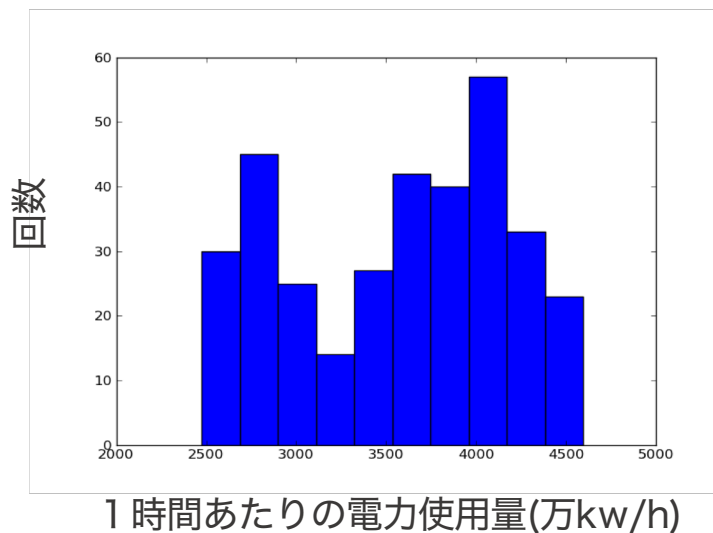
■ 指数分布

- 時間と共に回数が減る場合に見られる分布。
- 待ち時間、製品の故障、寿命などはこの分布になる。



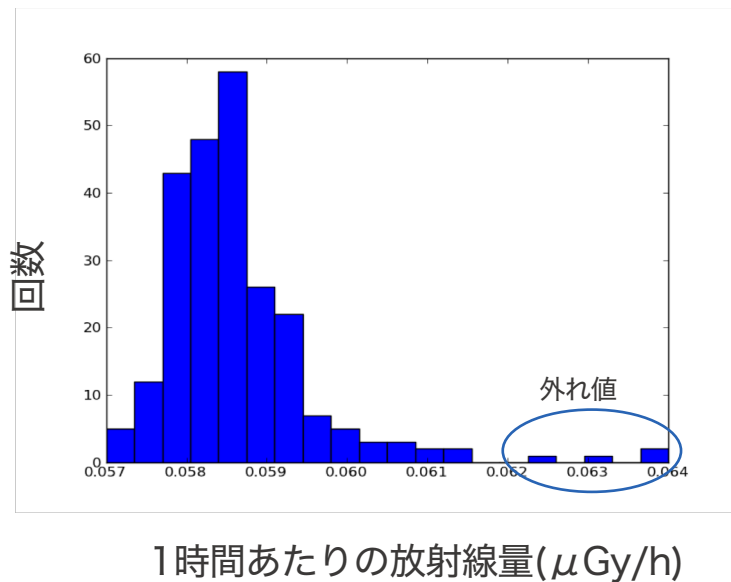
■ 例: 7月の1時間あたりの電力使用料

- ふたこぶ型の分布になっている
- 原因は夜と昼の電力使用量の性質が異なるためである。
- 昼と夜で層別が必要

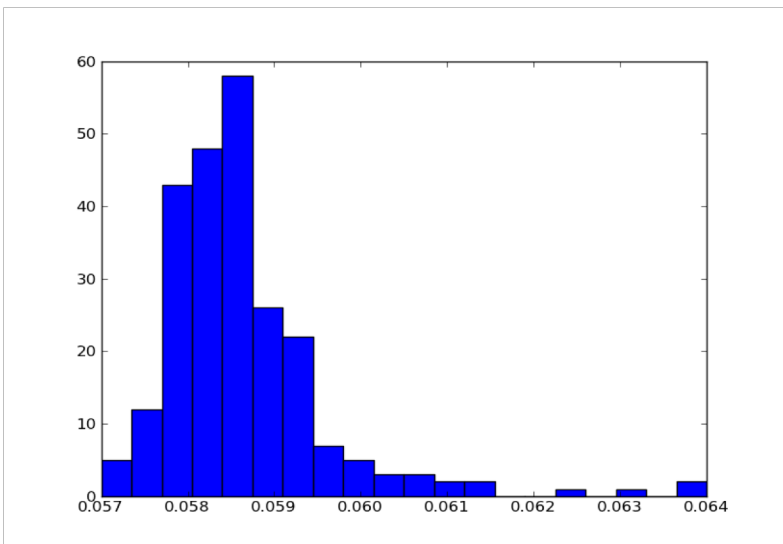
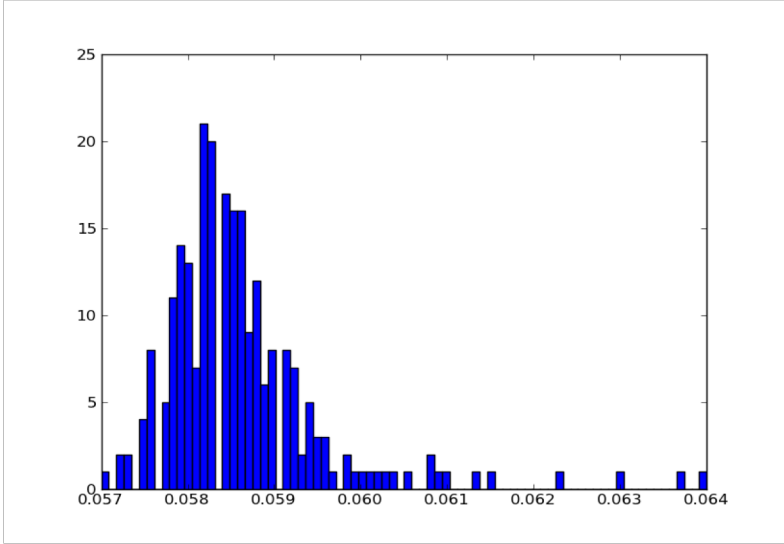


■ 例: 7月の新宿で観測された放射線量

- 基本的には正規分布ではあるが、外れ値が幾つか見られる。
- 外れ値がなぜ起こったか究明することが必要。



■ 作り方の悪いヒストグラム



グラフが歯抜けしているので良くない。
区間の設定が間違っている。

■ 分散、標準偏差

- 分布のばらつき具合を表す指標
 - 分散

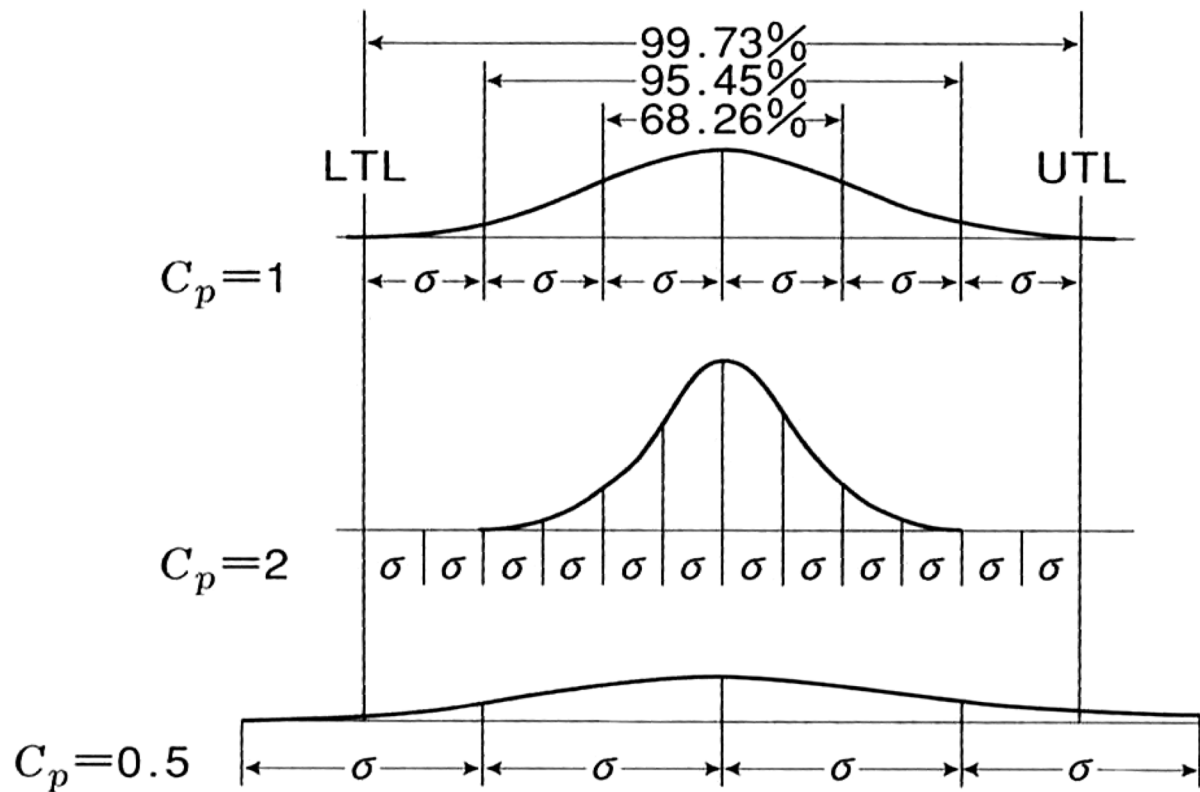
$$V = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x})^2$$

- 標準偏差

$$S = \sqrt{V}$$

(σ)

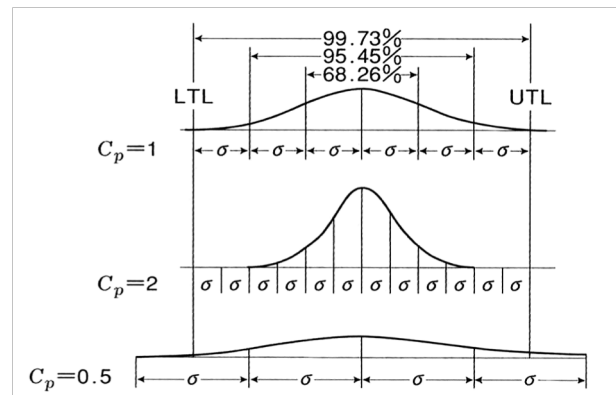
■ 標準偏差と分布の関係



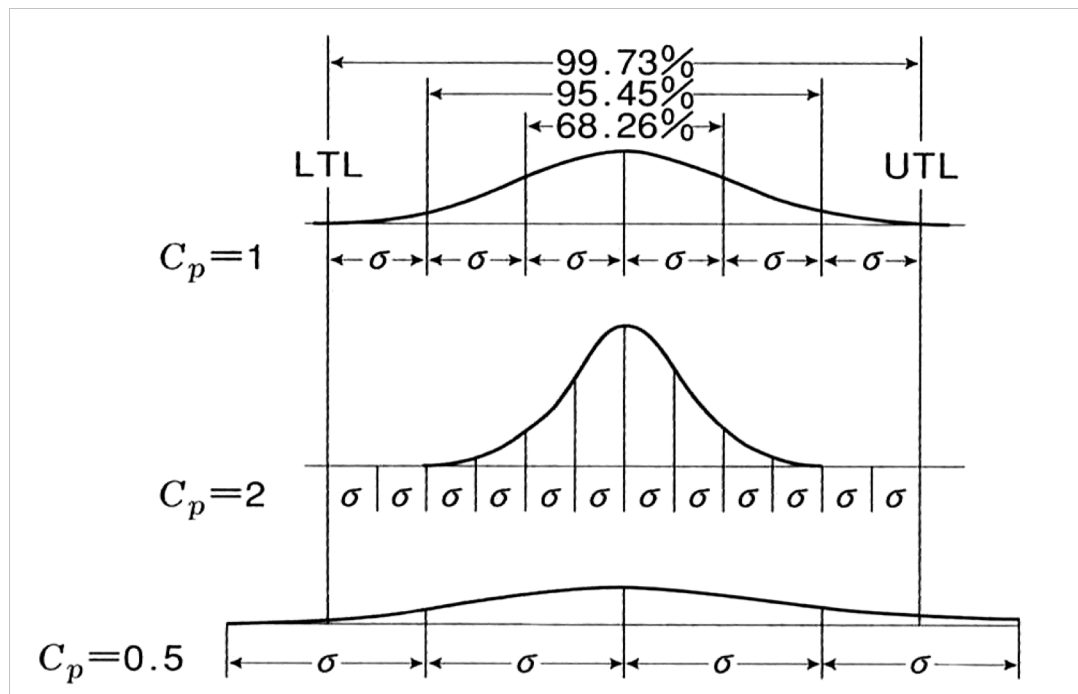
(QC数学の話 大村平より)

■ 工程能力指数

- 製品規格と分布の関係を表す指標
- 大きければ大きいほどよいが、大きすぎる場合過剰に対策をしている場合もある
- 上方許容限界(UTL)
 - 品質の上方限界
- 下方許容限界(LTL)
 - 品質の下方限界
- 工程能力指数



$$C_p = \frac{UTL - LTL}{6\sigma}$$



(QC数学の話 大村平より)

工程能力指数は製品の分布が規格内にどれくらい収まっているかという指標

$C_p=1$ なら生産された全製品のうち99.73%が規格内に収まっている。

$C_p > 1.67$	規格に対して余裕がありすぎ
$1.67 > C_p > 1.33$	工程能力は十分
$1.33 > C_p > 1.00$	規格に対して余裕が無い
$1.00 > C_p > 0.67$	工程能力が不足
$0.67 > C_p$	放置したら大変