

# サンプルサイズ及び データの処理

2009年6月3日

加藤洋一

# 必要なサンプルサイズ

- 事前に、最悪であるデータを必ずしも予想し、それを測定する、とは限らない と考えて、

条件：大きいほうの 10% であるデータが、少なくとも90%の確率で、サンプルに含まれる。

設計の考え方：サンプルサイズ $n$ のとき、その中がすべて確率0.9の大きくないものだけになる確率を10%以下にすればよい。

# 計算

$$0.9^{22} = 0.09847709$$

となる. この条件では,  $n=22$  となる.

## 【参考1】

- 含まれる確率を 95%, 99% とすると,

$$0.9^{29} = 0.047101287 \text{ より } n = 29$$

$$0.9^{44} = 0.009697737 \text{ より } n = 44$$

# 計算

【参考2】母集団のサイズが小さいとき,

- 超幾何分布で計算

$$N=50 \text{ のとき, } n = 18$$

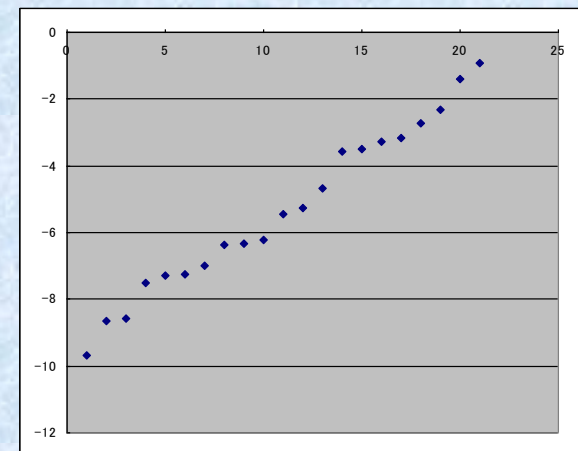
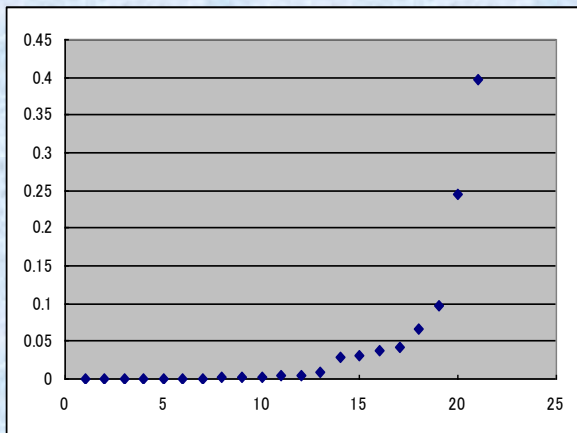
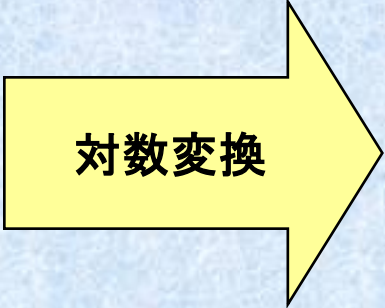
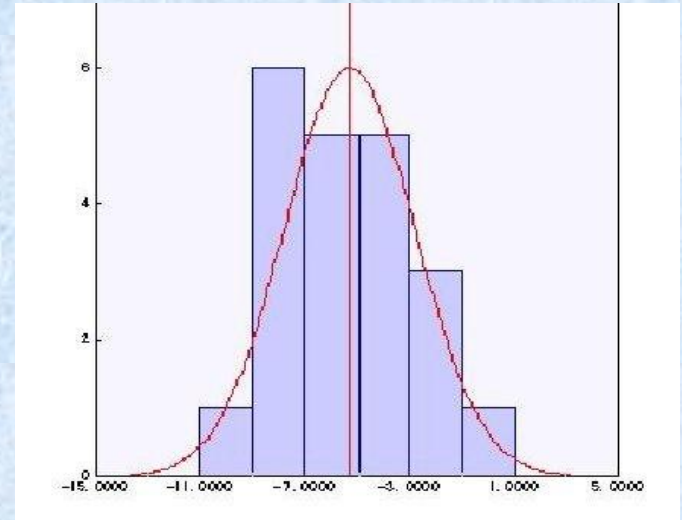
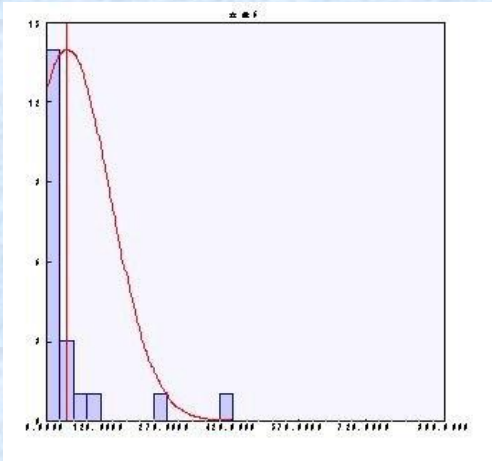
$$N=30 \text{ のとき, } n = 16$$

$$N=20 \text{ のとき, } n = 14$$

## 2. データの処理方法 (最悪値の推測)

- ① データを対数変換する.
  - ② データから最悪値を計算
    - A. データの中の最悪値
    - B. 平均値, 標準偏差 → データの区間推定 上側
    - C. 上位 10 のデータで, 平均値, 標準偏差  
→ 区間推定 上側
- <この中の大きな値を候補とする>
- ③ 値を元に戻す.

# データ例: コノバルト (n=21)



# データ例:コバルト(n=21)

A:データの最悪値:0.39640625

B:対数変換データで, 区間推定

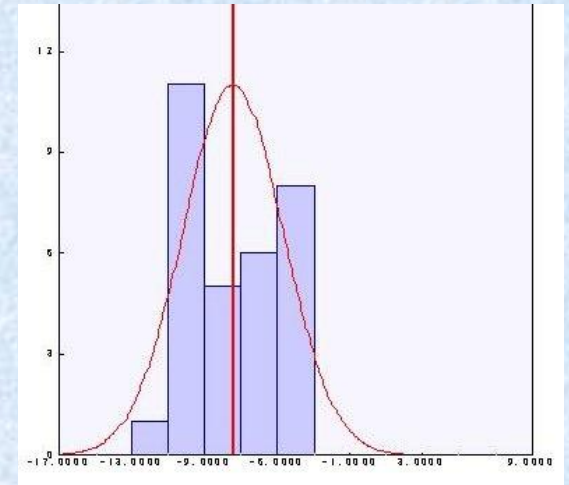
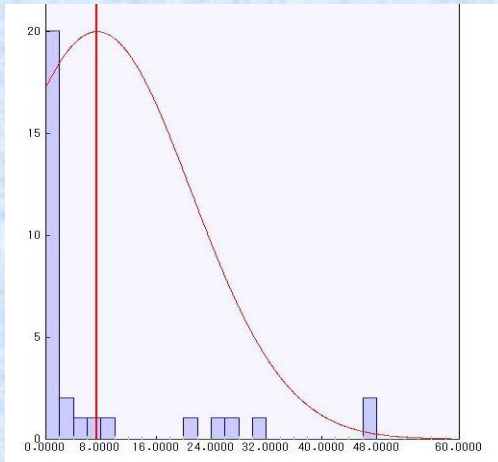
信頼率 上側90%:0.414821676



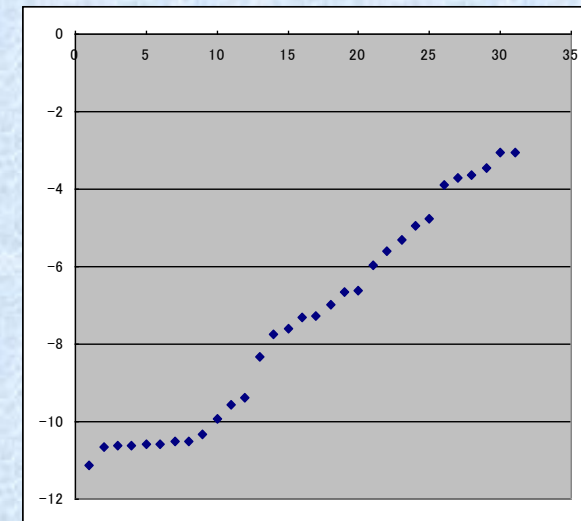
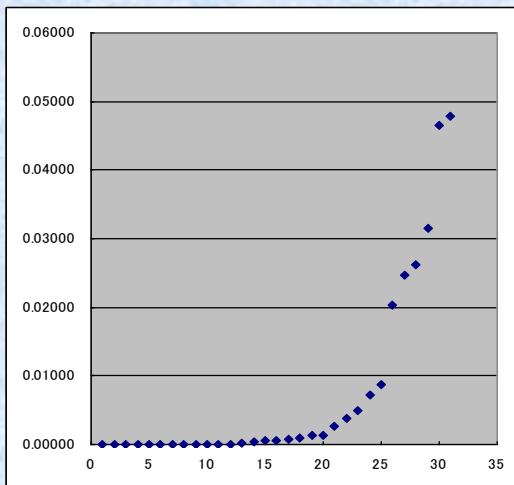
C:対数変換データの上位10で区間推定

信頼率 上側90%:0.315029518

# データ例：ヒドラジン (n=31)



対数変換





# データ例: ヒドラジン (n=31)

A: データの最悪値: 0.0479

B: 対数変換データで, 区間推定

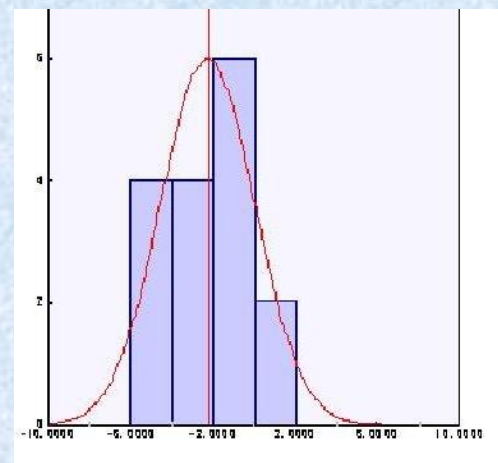
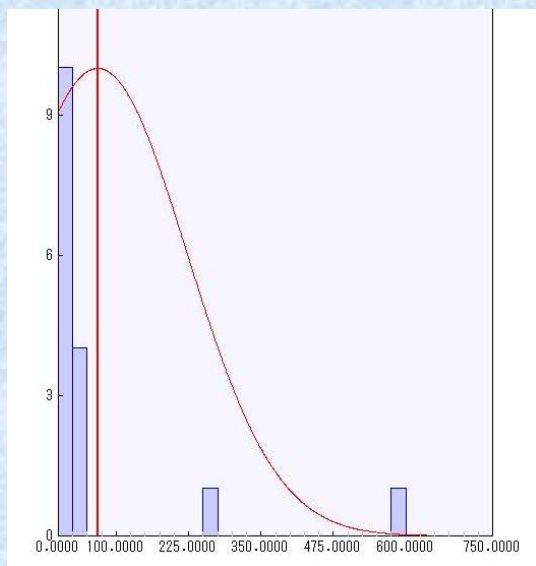
信頼率 上側90%: 0.076160985



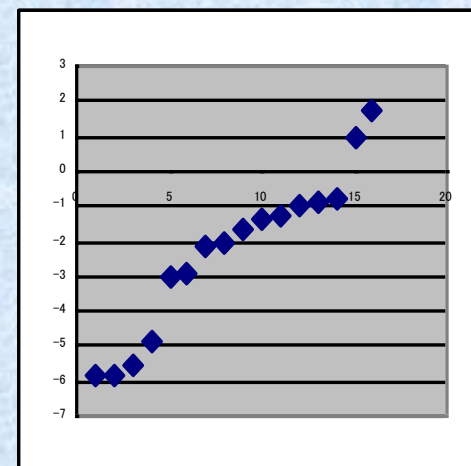
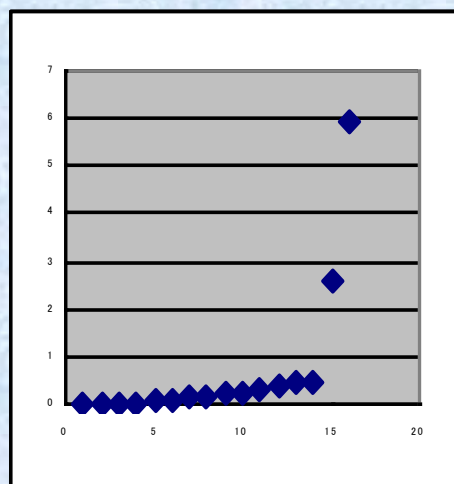
C: 対数変換データの上位10で区間推定

信頼率 上側90%: 0.053076567

# データ例：酸化プロピレン (n=16)



対数変換



# データ例：酸化プロピレン (n=16)

A: データの最悪値: 5.949

B: 対数変換データで, 区間推定

信頼率 上側90%: 6.41798



C: 対数変換データの上位10で区間推定

信頼率 上側90%: 2.719491