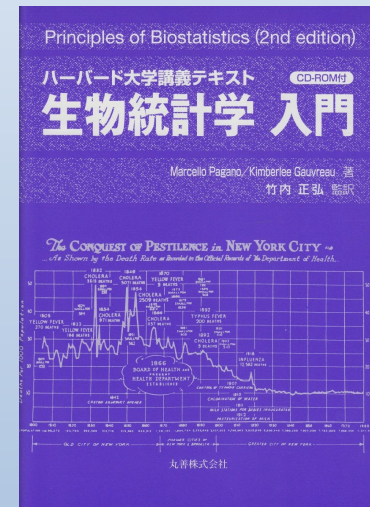


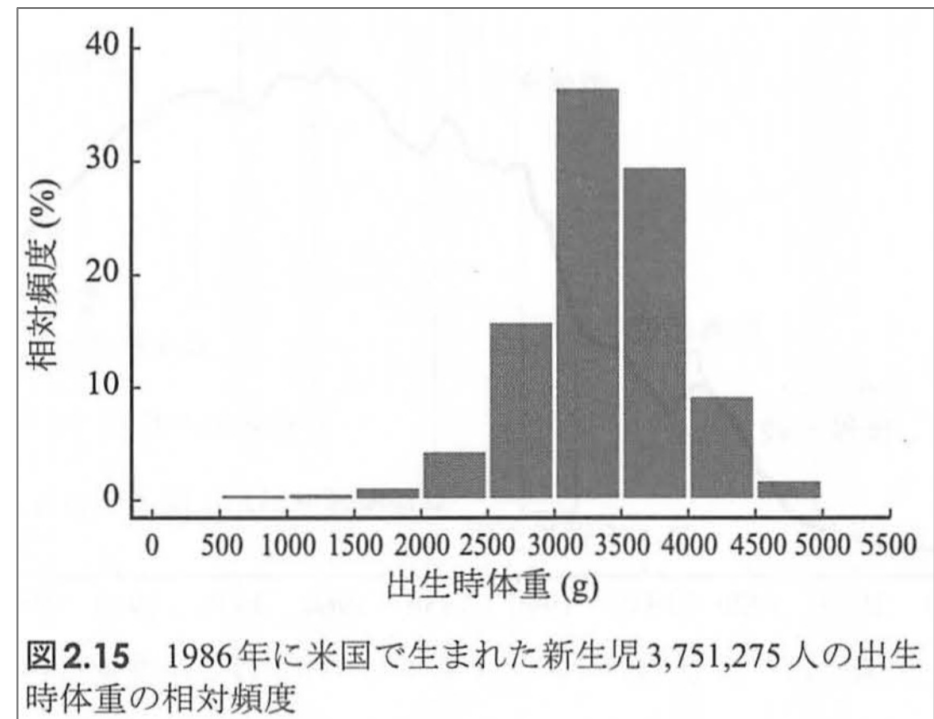
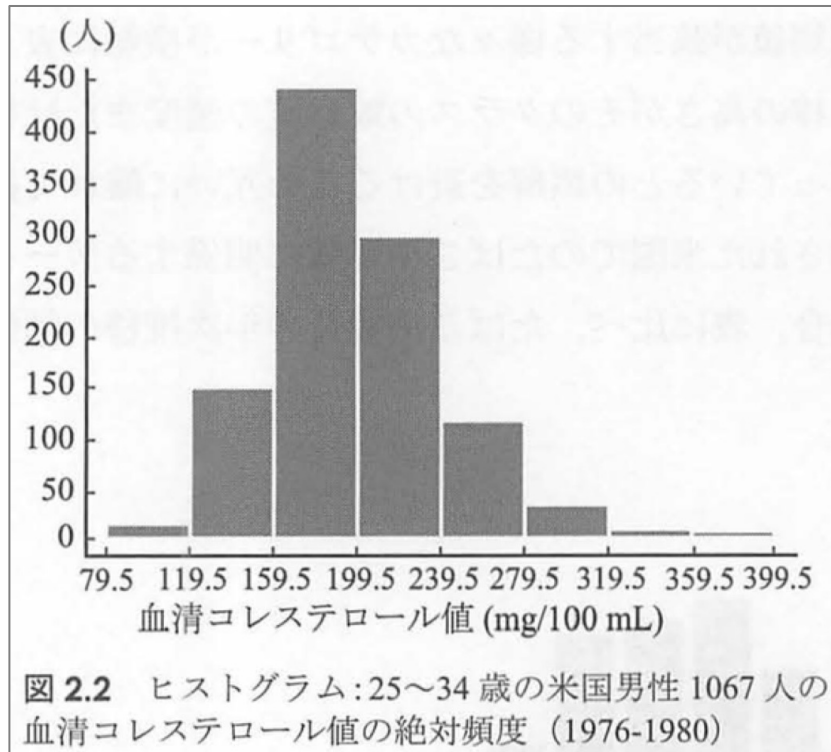
# 「データの表示」

参考書:

「ハーバード大学講義テキスト 生物統計学入門」2章  
(1章は序論なので飛ばします)



# 頻度は貴重な情報1:ヒストグラム



# 頻度は貴重な情報2: 箱ひげ図

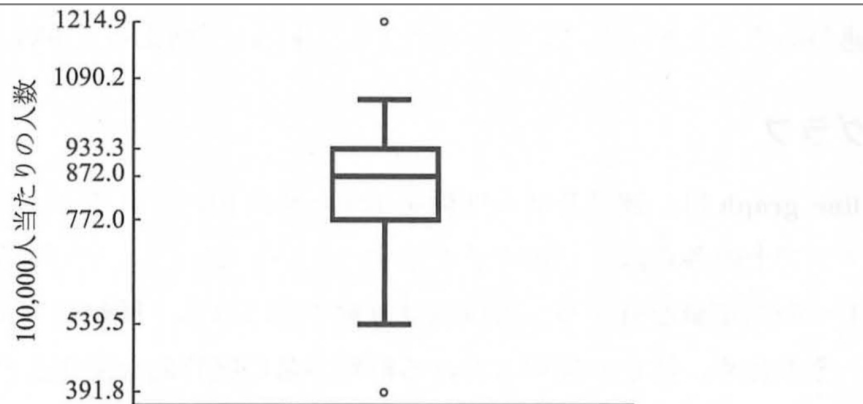


図 2.8 ボックスプロット：米国における粗死亡率（1992）

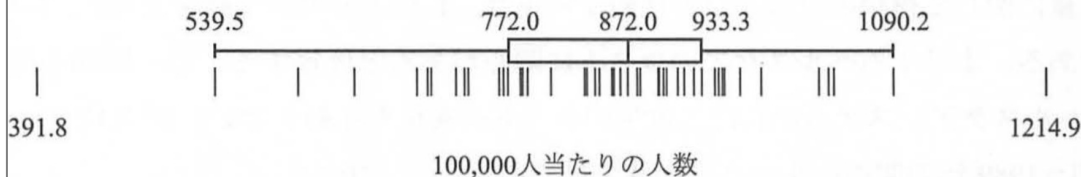


図 2.9 一元配置散布図およびボックスプロット：米国における粗死亡率（1992）

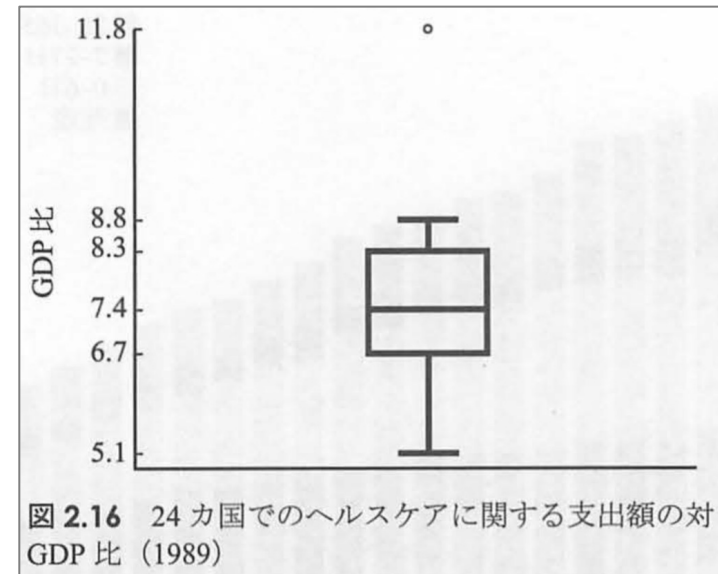


図 2.16 24 カ国でのヘルスケアに関する支出額の対 GDP 比（1989）

# 頻度は貴重な情報3: 離散データ

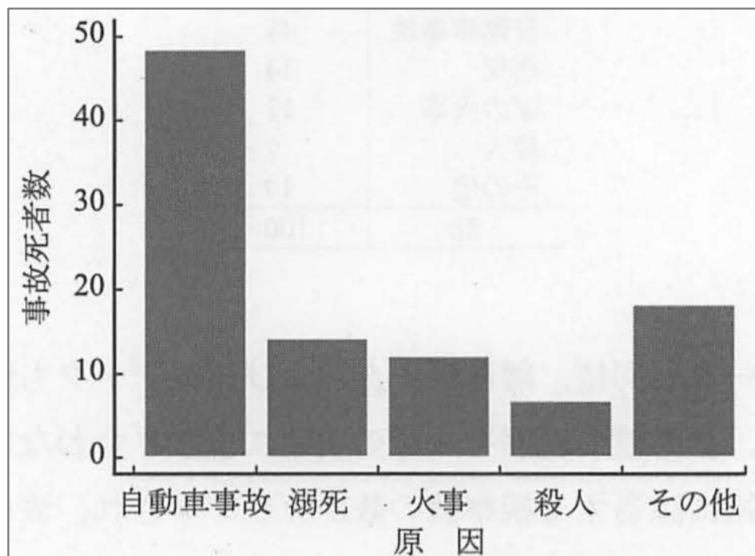


図 2.13 米国における 5～9 歳の子供 100 人の事故死 (1980-1985)

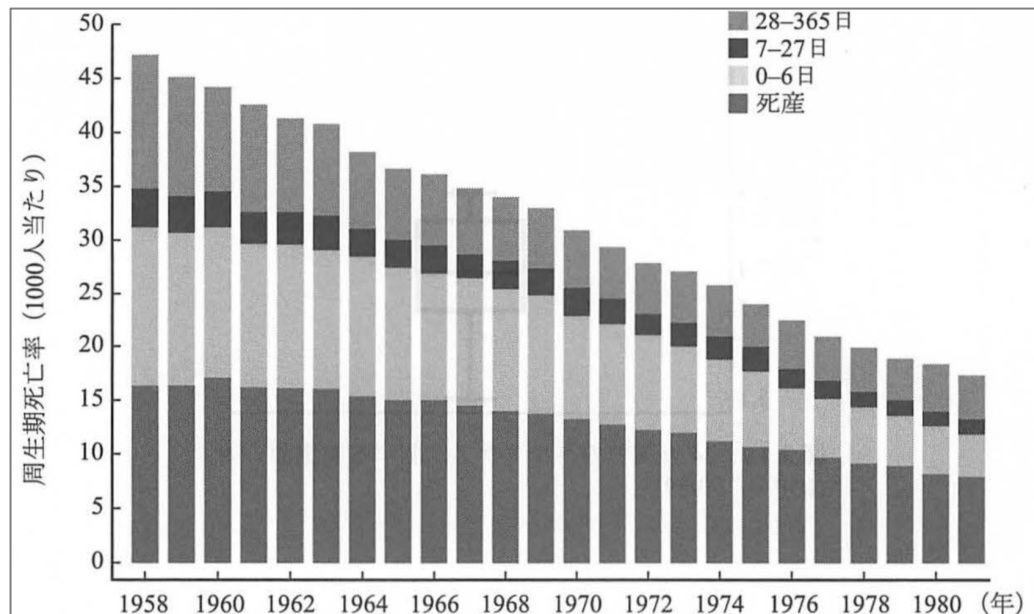
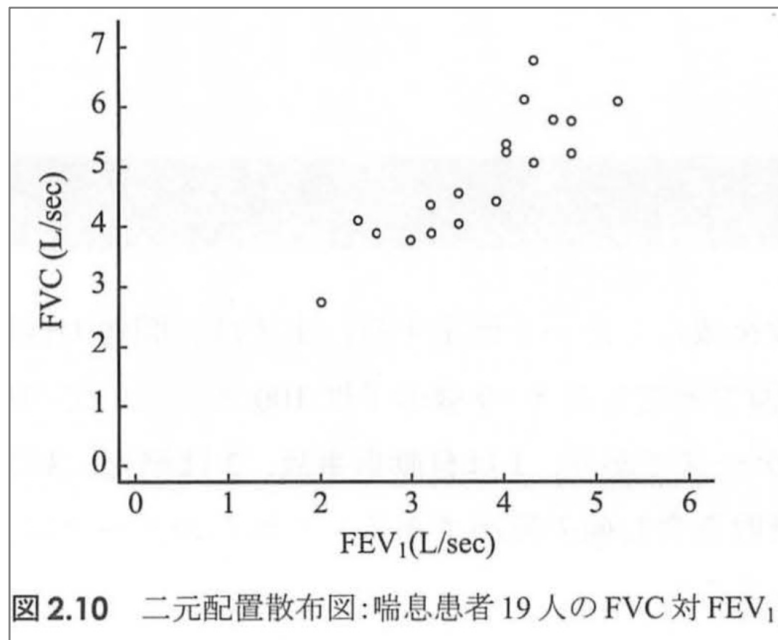


図 2.14 フランスにおける乳児および周生期死亡 (1978-1981)

# 2つの変数の関連を捉える



# 「数字による要約尺度」 (3章)

- 平均、分散、標準偏差、中央値などが手計算できるようになることが目標です。

# 平均と分散の計算法(喘息患者肺活量)

被験者	$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	2.30	-0.65	0.4225
2	2.15	-0.80	0.6400
3	3.50	0.55	0.3025
4	2.60	-0.35	0.1225
5	2.75	-0.20	0.0400
6	2.82	-0.13	0.0169
7	4.05	1.10	1.2100
8	2.25	-0.70	0.4900
9	2.68	-0.27	0.0729
10	3.00	0.05	0.0025
11	4.02	1.07	1.1449
12	2.85	-0.10	0.0100
13	3.38	0.43	0.1849
計	38.35	0.00	4.6596

$x$ : 最初の1秒間に肺から排出できる空気の体積

$$\begin{aligned}\bar{x} &= \frac{1}{13} \sum_{i=1}^{13} x_i && \text{平均(mean)} \\ &= \left( \frac{1}{13} \right) (2.30 + 2.15 + 3.50 + 2.60 + 2.75 + 2.82 + 4.05 + 2.25 + 2.68 + 3.00 + 4.02 + 2.85 + 3.38) \\ &= \frac{38.35}{13} \\ &= 2.95 \text{ リットル}\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{1}{(13-1)} \sum_{i=1}^{13} (x_i - 2.95)^2 && \text{分散(variance)} \\ &= \frac{4.6596}{12} \\ &= 0.39 \text{ リットル}^2\end{aligned}$$

$$\begin{aligned}s &= \sqrt{s^2} && \text{標準偏差(standard deviation)} \\ &= \sqrt{0.39} \text{ リットル}^2 \\ &= 0.62 \text{ リットル}\end{aligned}$$

小さい順に並べ替え:

2.15, 2.25, 2.30, 2.60, 2.68, 2.75, 2.82, 2.85, 3.00, 3.38, 3.50, 4.02, 4.05  
中央値(median)

# 平均と比べて中央値は頑健である

表 3.5 呼吸停止の状態にある喘息患者 10 人の心拍数

患者	心拍数(毎分)
1	167
2	150
3	125
4	120
5	150
6	150
7	40 ← 異常値
8	136
9	120
10	150

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i && \text{平均(異常値も含む)} \\ &= \left(\frac{1}{10}\right)(167 + 150 + 125 + 120 + 150 + 150 + 40 + 136 + 120 + 150) \\ &= \frac{1308}{10} \\ &= 130.8 \text{ 回/分}\end{aligned}$$

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i && \text{平均(異常値を除く)} \\ &= \left(\frac{1}{9}\right)(167 + 150 + 125 + 120 + 150 + 150 + 136 + 120 + 150) \\ &= \frac{1268}{9} \\ &= 140.9 \text{ 回/分}\end{aligned}$$

小さい順に並べ替え:

40, 120, 120, 125, 136, 150, 150, 150, 150, 167

$$(136 + 150) / 2 = 143 \text{ 回}$$

中央値(異常値も含む)



# 例題：試験の得点(70, 70, 80, 90, 90)の、平均、分散、標準偏差、中央値を求めよ

$$[\text{平均}] = (70+70+80+90+90)/5 = \underline{80}$$

$$\text{得点} - [\text{平均}] = -10, -10, 0, 10, 10.$$

$$\begin{aligned} [\text{分散}] &= \{ (-10)^2 + (-10)^2 + 0 + 10^2 + 10^2 \} / (5 - 1) \\ &= \{ 100 + 100 + 0 + 100 + 100 \} / 4 \\ &= \underline{100} \end{aligned}$$

$$[\text{標準偏差}] = \text{分散のルート} = \sqrt{100} = \underline{10}$$

平均±標準偏差は、 $80 \pm 10$ と書ける。

中央値は、3番目に小さい値であるので、80である。