クレジット:
UTokyo Online Education 統計データ解析 I 2017 小池祐太

東京大学 THE UNIVERSITY OF TOKYO

CC BY NC ND

統計データ解析Ｉ (平成２９年度)

東京大学大学院数理科学研究科
統計データ解析教育研究グループ

村田　昇 (早稲田大学, 東京大学)
吉田朋広 (東京大学)
小池祐太 (首都大学東京, 東京大学)

第 7 章 基礎的な記述統計量とデータの集約

記述統計量とはデータを簡潔に要約して表すための統計値のことで, 要約統計量, 基本統計量とも言われる. ヒストグラム (あるいは密度関数) や箱ひげ図などのグラフと併用して, その集団全体の特徴を表す重要な指標となる. 本章では, 比較的良く用いられる統計量を, その背景となるモーメント, 順序, 分布という考え方に基づいて分類する.

7.1. モーメントに基づく統計量

7.1.1. 平均・分散・標準偏差. n 個のデータ $X_1, X_2, \ldots, X_n$ が与えられたとき, それらを代表する値として, **(標本) 平均**

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

が頻繁に利用される. また, データのばらつき具合の指標として, **(標本) 分散**

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

およびその平方根である **(標本) 標準偏差**が広く利用されている.

5–6 章で述べた通り, 確率統計学では, 一つ一つのデータ X_i をある確率変数の実現値とみなすことで, データの背後にある現象に対する統計解析を行う. 確率変数 $X_1, X_2, \ldots, X_n$ が同分布であれば, 5–6 章で定義したように, X_i たちに共通の平均 μ および分散 σ^2 を考えることができる (もちろん適切な次数のモーメントの存在を仮定した下で). さらに, $X_1, X_2, \ldots, X_n$ が独立であれば, 大数の強法則より, 標本平均・標本分散・標本標準偏差はそれぞれ $n \to \infty$ のとき確率 1 で平均 μ・分散 σ^2・標準偏差 σ に収束する. これは, 標本平均・標本分散・標本標準偏差をそれぞれサンプル対象の集団の「真の」平均・分散・標準偏差の推定量と考えた場合に, これらの推定量がサンプル数 n が十分大のときに「まともな」推定量であるという根拠の 1 つを与える. このような性質を推定量の **(強) 一致性**と呼び, 一致性をもつ推定量を **(強) 一致推定量**と呼ぶ.

一致性はサンプル数が十分大きい場合に推定量がまともであることの 1 つの根拠を与えるが, サンプル数が小さい場合の推定量の性質については何も語っていない. そのような場合の推定量の良さに関する性質の 1 つとして**不偏性**がある. 一般にパラメーター θ の推定量 $\hat{\theta}$ が不偏であるとは, $\hat{\theta}$ の平均が θ, すなわち

$$E[\hat{\theta}] = \theta$$

が成り立つことをいう. 標本平均は μ の不偏推定量である. すなわち,

$$E[\bar{X}] = \mu$$

が成り立つ. 一方で, 標本分散は σ^2 の不偏推定量ではない. 実際,

$$E[S^2] = \frac{n-1}{n}\sigma^2$$

が成り立つ. この式は, 標本分散は平均的には真の分散を過小推定する傾向にあることを意味する. このバイアスを補正するには, 標本分散に $n/(n-1)$ をかけてやれば良い. すなわち,

$$s^2 := \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

は σ^2 の不偏推定量となる. わざわざ不偏性を持たない S^2 を σ^2 の推定量として使う理由は通常ないので, 標本分散という場合には s^2 のことを指す場合もあるが, バイアス補正をしていることを強調するために**不偏分散**と呼ぶ場合もある. R には不偏分散を計算するための関数として `var()` が用意されている. 同様に, 標本標準偏差という場合は, 通常, 不偏分散の平方根 s を指し, R では関数 `sd()` で計算できる (ただし, 一般に s は標準偏差 σ の不偏推定量ではない).

```
> ## 標本分散が平均的には過小推定となることの確認
> set.seed(123) # 乱数の初期値を指定
> sample.var <- function(n){ # n 個の標準正規乱数の標本分散を計算する関数
+   x <- rnorm(n)
+   return(mean((x - mean(x))^2))
+ }
> n <- 10 # サンプル数
> MC <- 10000 # 実験回数
> v <- replicate(MC, sample.var(n)) # sample.var(n) を MC 回実行して結果を記録
> mean(v) # 1-1/n=0.9 に近い (真の分散 1 を過小推定している)
[1] 0.9009275
> mean((n/(n-1))*v) # バイアス修正すると 1 に近くなる
[1] 1.001031
```

(unbiased.r)

複数のデータを同時に分析する場合, 単位や基準を揃えた方が扱いやすい. このような目的でよく使われる方法に, データの**標準化**がある. データ $X_1, X_2, \ldots, X_n$ の標準化は

$$Z_i = \frac{X_i - \bar{X}}{s} \qquad (i = 1, 2, \ldots, n)$$

で定義される.[1] 定義から明らかなように, $Z_1, Z_2, \ldots, Z_n$ の標本平均は 0, 不偏分散は 1 である (むしろ, そうなるようにデータを一次変換したものが標準化である). 標準化は**標準得点**あるいは **Z スコア**とも呼ばれる. 一方で, 教育学や心理学では, データを標本平均が 50, 標準偏差 (不偏分散の平方根) が 10 となるように一次変換したもの

$$T_i = 10Z_i + 50 \qquad (i = 1, \ldots, n)$$

を使う場合が多い. これを**偏差値得点**あるいは **T 得点**と呼ぶ.

```
> ## 気候データによる例
> kikou <- read.csv("kikou2016.csv", fileEncoding = "sjis") # データの読み込み
> dat <- subset(kikou, select = -c(月, 日)) # 月日は計算対象から削除
> head(dat)
  気温 降水量 日射量 風速
1  7.5      0  11.80  2.6
2  7.3      0  11.59  1.9
3  9.3      0  10.77  1.4
4  9.2      0  11.19  1.6
5 10.9      0  10.57  1.8
6  8.9      0   4.54  1.9
> dat.std <- scale(dat) # データを各変数ごとに標準化
> head(dat.std)
```

[1] s の代わりに S で割って定義する文献もある.

```
        気温       降水量       日射量         風速
1 -1.1682298 -0.3583779 -0.1233405 -0.2189065
2 -1.1942766 -0.3583779 -0.1527083 -1.0461017
3 -0.9338081 -0.3583779 -0.2673829 -1.6369554
4 -0.9468315 -0.3583779 -0.2086471 -1.4006139
5 -0.7254333 -0.3583779 -0.2953523 -1.1642724
6 -0.9859018 -0.3583779 -1.1386296 -1.0461017
> colMeans(dat.std) # 各変数の平均が 0 であることの確認
          気温          降水量          日射量          風速
-1.460959e-16  1.178285e-17 -6.278182e-17 -9.858538e-18
> apply(dat.std, 2, "sd") # 各変数の標準偏差が 1 であることの確認
  気温 降水量 日射量   風速
     1      1      1      1
```

(scale.r)

7.1.2. 歪度と尖度. 中心極限定理が示すように, 正規分布は確率分布のうち最も基本的なものと考えられる. 正規分布は平均と分散を決めると完全に決定されるから, 正規分布に従うデータを考える際には標本平均と標本分散 (不偏分散) を考えれば十分である. しかし, 現実には正規分布では捉えきれない特徴をもつデータに遭遇することがしばしばある. そのようなデータを考える場合の最初のアプローチとして, 正規分布からのずれを調べるということがしばしば行われる. そのための統計量として代表的なものに歪度と尖度がある.

X を平均 μ, 分散 σ^2 をもつ確率変数とする. X が 3 次のモーメントをもつとき,

$$\frac{E[(X-\mu)^3]}{\sigma^3}$$

を**歪度**と呼ぶ. 歪度は分布の非対称性を表す統計量で, 正の場合分布の右の裾の方が重く, 負の場合分布の左の裾の方が重いと考えられる. 左右に対称的な分布の歪度は 0 であり, 従って正規分布の歪度は 0 である. 正の歪度をもつ分布としては, 例えばガンマ分布 $\Gamma(\nu,\alpha)$ があり, その歪度は $2/\sqrt{\nu}$ で与えられる.

一方で, X が 4 次のモーメントをもつとき,

$$\frac{E[(X-\mu)^4]}{\sigma^4}$$

を**尖度**と呼ぶ. 尖度は平均の周囲の分布の尖り具合を表す統計量だと考えられる. 正規分布の場合 3 であるため, 正規分布との比較のため上の定義から 3 を引いた量

$$\frac{E[(X-\mu)^4]}{\sigma^4} - 3$$

のことを尖度と呼ぶ文献も多いが, 後者を前者と区別するために**超過尖度**と呼ぶ場合もある. 超過尖度が正の分布は正規分布よりも平均の周囲の分布が尖っており, 負の分布は丸みを帯びていると考えられる. 前者の場合, 平均まわりの密度が分布の裾の方にまわっていることが多いため, 正規分布より裾が重いと解釈されることが多い. 正の超過尖度をもつ分布としては, 例えば自由度 $\nu > 4$ をもつ t 分布 $t(\nu)$ があり, その超過尖度は $6/(\nu-4)$ で与えられる ($\nu \leq 4$ のときは $t(\nu)$ は 4 次モーメントをもたない). また, ガンマ分布 $\Gamma(\nu,\alpha)$ は超過尖度 $6/\nu$ をもつ.

観測データ $X_1, X_2, \ldots, X_n$ から歪度と尖度を推定するには, それらの標本バージョンを考えればよい. すなわち, 歪度の推定量としては, **標本歪度**

$$\frac{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^3}{s^3}$$

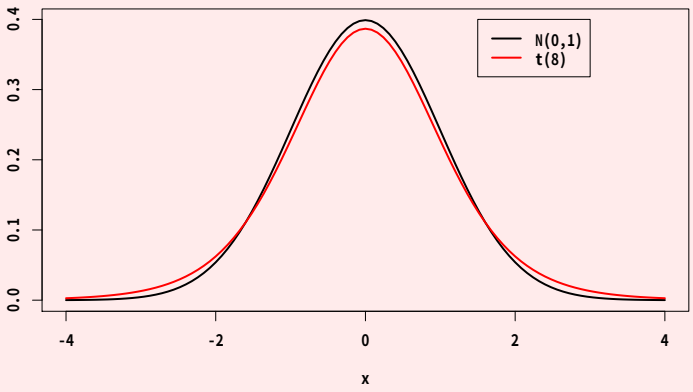
を考えればよく, 尖度の推定量としては, **標本尖度**

$$\frac{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^4}{s^4}$$

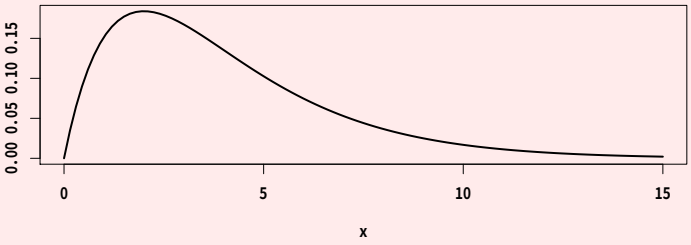
を考えればよい. 標本歪度・標本尖度を計算するための関数はデフォルトでは R には実装されていないため, 自作するかパッケージを利用する. 例えば, パッケージ e1071 には標本歪度を計算するための関数 skewness() および標本尖度を計算するための関数 kurtosis() が実装されている. 後者は上の定義の標本尖度から 3 を引いたもの (すなわち標本超過尖度) を計算することに注意せよ.

標本歪度・標本尖度の値は標本平均・分散に比べてばらつきが大きくなる傾向があるため, サンプル数が少ない場合の計算結果の解釈には注意を要する.

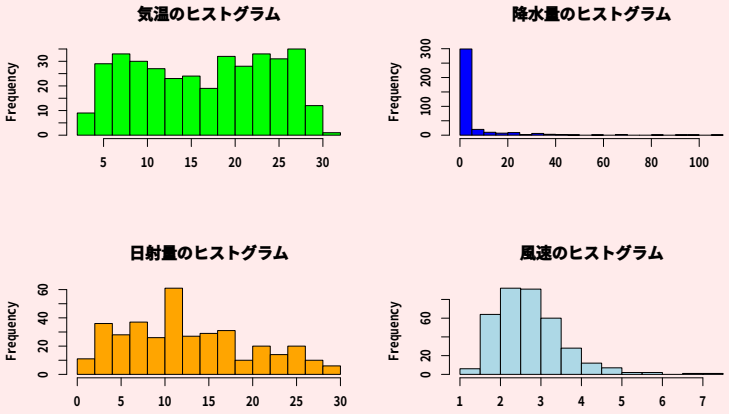
```
> install.packages("e1071") # パッケージのインストール
> library(e1071) # パッケージのロード
> ## シミュレーションによる例
> set.seed(123)
> x <- rnorm(10000) # 標準正規乱数を 10000 個発生
> skewness(x) # 歪度：0 に近い
[1] 0.008197965
> kurtosis(x) # (超過) 尖度：0 に近い
[1] 0.01073845
> y <- rt(10000, df = 8) # 自由度 8 の t 乱数を 10000 個発生
> skewness(y) # 歪度：0 に近い
[1] 0.01196181
> kurtosis(y) # (超過) 尖度：6/(8-4)=1.5 に近い
[1] 1.445932
> # グラフで確認してみる
> plot(dnorm, -4 ,4 , lwd = 2, ylab = "") # 標準正規密度のプロット
> curve(dt(x, df = 8), add = TRUE, lwd = 2, col = "red") # 自由度 8 の t 分布の密度のプロット
> legend(1.5, 0.4, legend = c("N(0,1)","t(8)"),
+        lwd = 2, col = c("black", "red")) # 凡例の追加
```




```
> z <- rchisq(10000, df = 4) # 自由度 4 のカイ二乗乱数を 10000 個発生
> skewness(z) # 歪度：sqrt(8/4)=1.414... に近い
[1] 1.415016
> kurtosis(z) # (超過) 尖度：12/4=3 に近い
[1] 2.973832
> # グラフで確認してみる
> curve(dchisq(x,df=4),0,15,lwd=2,ylab="") # 自由度 4 のカイ二乗分布の密度のプロット
```




```
> ## 気候データによる例
> kikou <- read.csv("kikou2016.csv", fileEncoding = "sjis") # データの読み込み
> dat <- subset(kikou, select = -c(月, 日)) # 月日は計算対象から削除
> apply(dat, 2, "skewness") # 歪度
      気温       降水量       日射量         風速
-0.0508104  4.4538316  0.4333552  1.3670592
> apply(dat, 2, "kurtosis") # (超過) 尖度
      気温       降水量       日射量         風速
-1.3089167 23.4751756 -0.6806749  3.5644397
> # ヒストグラムによる確認
> op <- par(mfrow = c(2,2))
> cl <- c("green", "blue", "orange", "lightblue") # 色を用意
> nam <- colnames(dat) # 変数名
> for(i in 1:4){
+   hist(dat[ ,nam[i]], col = cl[i], breaks = 20, xlab = "",
+        main = paste0(nam[i], "のヒストグラム"))
+ }
> par(op)
```



(skewkurt.r)

演習 7.1. 正規分布からサンプルされたデータから計算された標本歪度・標本尖度の不偏性について調べてみよ.

7.1.3. 相関と共分散. 複数のデータが与えられた場合, それらのデータの間の関係性を知りたい場合が頻繁に生じる. そのような目的のための最も基本的な記述統計量に **(標本) 相関**があり, これは 2 種類のデータ間の比例関係の大きさを計測する. 2

種類のデータ $x_1, x_2, \ldots, x_n$ および $y_1, y_2, \ldots, y_n$ に対して, それらの相関は

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{7.1}$$

で定義される. ここに, $\bar{x}$ および $\bar{y}$ はそれぞれ $x_1, x_2, \ldots, x_n$ および $y_1, y_2, \ldots, y_n$ の標本平均である. 相関は -1 以上 1 以下の値をとり, 1 に近いほど正の比例関係が強く, -1 に近いほど負の比例関係が強いことになる. なお, (7.1) の分子の統計量を $n-1$ で割ったものは **(標本) 共分散**と呼ばれる. 相関および共分散はそれぞれ関数 `cor()` および関数 `cov()` で計算できる. 2 種類のデータ `x` および `y` が与えられたとき, それらの相関は `cor(x,y)` で計算できる. 一方で, `x` がデータフレームのとき, `cor(x)` は (i, j) 成分が `x` の i 列と j 列の間の相関であるような行列 (**相関行列**) を計算する. 共分散についても同様である.

```
> ### sleep データによる例
> ### 2 種類の睡眠薬の効果の個人差に相関はあるか?
> x <- subset(sleep, group == 1, extra)
> y <- subset(sleep, group == 2, extra)
> cor(x, y)
          extra
extra 0.7951702
> ## 気候データによる例
> kikou <- read.csv("kikou2016.csv", fileEncoding = "sjis")
> cor(kikou[ ,-c(1:2)])
             気温       降水量       日射量       風速
気温    1.00000000  0.08575259  0.3314077 0.14291472
降水量 0.08575259  1.00000000 -0.3607787 0.07572892
日射量 0.33140773 -0.36077872  1.0000000 0.31826364
風速    0.14291472  0.07572892  0.3182636 1.00000000
```

(correlation.r)

演習 7.2. R の組込データセット `state.x77` について, 列ごとの標本平均, 標準偏差, 標本歪度, 標本尖度を求めよ. また, 相関行列も求めよ.

7.2. 順序に基づく統計量

データの順序にもとづく記述統計量もよく利用される. 例えば, $X_1, \ldots, X_n$ の**最大値**は関数 `max()` で, **最小値**は関数 `min()` でそれぞれ計算できる. データを

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

のように昇順に並べ替えた際に中央の位置にくる値を**メディアン**もしくは**中央値**と呼ぶ. n が奇数の場合, メディアンは $X_{((n+1)/2)}$ であり, n が偶数の場合は $(X_{(n/2)} + X_{(n/2+1)})/2$ である. メディアンは関数 `median()` で計算できる.

メディアンは平均と同様データを代表する値だと考えられるが, 平均と比較して, 計算結果がデータに含まれる異常な値 (**外れ値**と呼ばれる) の影響を受けにくい.

メディアンの一般化として, $\alpha \in [0, 1]$ に対して, その点以下のデータの個数が全体の (約)$100\alpha\%$ になるような点を $100\alpha\%$**分位点**と呼ぶ. 特に 25%分位点および 75%分位点をそれぞれ**第 1 四分位点**, **第 3 四分位点**と呼ぶ. **第 2 四分位点**は 50%分位点となるが, これはメディアンのことである. ベクトル `x` の 100`alpha`%分位点は `quantile(x, alpha)` で計算できる. 分位点は一意的には定まらず, いくつかの計算方式がある: `help(quantile)` を参照すること.

```
> ## 気候データによる例
> kikou <- read.csv("kikou2016.csv", fileEncoding = "sjis") # データの読み込み
```

```
> x <- kikou$気温
> mean(x) # 平均
[1] 16.47022
> median(x) # メディアン
[1] 17.15
> quantile(x, 0.5) # 上と同じ
  50%
17.15
> quantile(x, c(0.25, 0.75)) #第 1 四分位点および第 3 四分位点
   25%    75%
 9.425 23.075
> fivenum(x) # 五数要約 (最小・最大値および四分位点)
[1]  2.80  9.40 17.15 23.10 31.90
> # 分位点の計算方式の違いのため, quantile の結果と多少異なることに注意
> y <- c(x, 1000) # データに外れ値を加えてみる
> mean(y) # mean(x) と大きく異なる
[1] 19.15014
> median(y) # median(x) とあまり変わらない
[1] 17.2
> summary(kikou[ ,-c(1, 2)]) # データフレームの各列の五数要約と平均を計算
      気温            降水量            日射量           風速
 Min.   : 2.800   Min.   :  0.000   Min.   : 1.11   Min.   :1.200
 1st Qu.: 9.425   1st Qu.:  0.000   1st Qu.: 7.16   1st Qu.:2.200
 Median :17.150   Median :  0.000   Median :11.69   Median :2.700
 Mean   :16.470   Mean   :  4.861   Mean   :12.68   Mean   :2.785
 3rd Qu.:23.075   3rd Qu.:  2.000   3rd Qu.:17.58   3rd Qu.:3.200
 Max.   :31.900   Max.   :106.500   Max.   :29.87   Max.   :7.200
                                                                (order.r)
```

確率分布に対しても分位点が定義され, 推定や検定において重要な役割を果たす. $0 < \alpha < 1$ とする. 連続分布の 100α%**分位点**は, その分布に従う確率変数を X としたとき, 不等式

$$P(X \leq x) \geq \alpha$$

を満たす実数 x のうち最小のものとして定義される.[2] そのような実数は常に存在し, それを q_α とすると,

$$P(X \leq q_\alpha) = \alpha$$

が成り立つことが知られている. $X_1, X_2, \ldots, X_n$ が独立同分布な確率変数の列のとき, $X_1, X_2, \ldots, X_n$ の $100\alpha\%$分位点は $n \to \infty$ のとき X_i たちの従う分布の $100\alpha\%$分位点の一致推定量となることが知られている.

確率分布の分位点は, その分布の省略形が `xxx` の場合 (6.2.2 節参照), 関数 `qxxx()` で計算できる. 例えば, 平均 `mu`, 標準偏差 `sigma` の正規分布の 100`alpha`%分位点は,

```
qnorm(alpha, mean = mu, sd = sigma)
```

で計算できる.

```
> # データから計算された分位点が確率分布の分位点の
> # 一致推定量となることの確認
> set.seed(123)
> x <- rnorm(1000) # 標準正規乱数を 1000 個発生
```

[2] より一般に, 確率分布の $100\alpha\%$分位点は, その分布に従う確率変数を X としたとき, 不等式

$$P(X \leq x) \geq \alpha$$

を満たす実数 x の下限として定義される (そのような実数が存在しない場合は ∞ とする).

```
> alpha <- c(0.25, 0.5, 0.75) # 計算する分位点の位置
> quantile(x, probs = alpha) # データから計算される分位点
        25%         50%         75%
-0.628324243  0.009209639  0.664601867
> qnorm(alpha) # 確率分布の分位点
[1] -0.6744898  0.0000000  0.6744898
```

(quantile.r)

順序に基づいてデータのばらつきを測るための記述統計量もいくつか存在する. そのようなものとして, 最大値と最小値の差である**範囲**がある. 範囲は外れ値の影響を大きく受けるので, 第 3 四分位点と第 1 四分位点の差である**四分位範囲**もよく使われる. また, データ $X_1, X_2, \ldots, X_n$ のメディアンを m としたとき, $|X_1 - m|, |X_2 - m|, \ldots, |X_n - m|$ のメディアンを**メディアン絶対偏差**と呼ぶ.

```
> ## 気候データによる例
> kikou <- read.csv("kikou2016.csv", fileEncoding = "sjis") # データの読み込み
> dat <- subset(kikou, select = -c(月, 日)) # 月日は計算対象から削除
> apply(dat, 2, "sd") # 変数ごとの標準偏差
      気温      降水量      日射量        風速
 7.6784711 13.5629323  7.1506723  0.8462331
> range(dat$気温) # 最小値と最大値を計算
[1]  2.8 31.9
> diff(range(dat$気温)) # 範囲を計算
[1] 29.1
> apply(dat, 2, function(x) diff(range(x))) # 変数ごとの範囲
  気温 降水量 日射量   風速
 29.10 106.50  28.76   6.00
> apply(dat, 2, "IQR") # 変数ごとの四分位範囲
   気温   降水量   日射量     風速
13.6500  2.0000 10.4225  1.0000
> apply(dat, 2, "mad", constant = 1) # 変数ごとのメディアン絶対偏差
  気温 降水量 日射量   風速
  6.75   0.00   5.28   0.50
```

(range.r)

演習 7.3. 順序に基づく記述統計量について調べてみよう.

(1) 正規乱数から計算された四分位範囲・メディアン絶対偏差と標準偏差の間の関係を調べてみよ.
(2) 順序に基づいて複数のデータの間の関係性を要約するための記述統計量について調べてみよ.

7.3. 頻度に基づく統計量

データの中で最も頻度が高く現れる値を, **モード**もしくは**最頻値**と呼ぶ. モードはデータが有限個の値を取る場合に特に有効であるが, データが連続で無限に多くの値を取ることができる場合には注意が必要である. 連続なデータの場合でも有限個の観測データに対してモードは定義できるが, 偶々観測値として現れた値なので, その意味はよく考えなくてはならない. 必要に応じて, 例えば区分的に集計するなどの工夫をすることもある.

```
> ## モードを計算するための関数の作成
> mode <- function(x){
```

```
+   obj <- table(x) # 度数分布表の作成
+   return(names(obj)[obj == max(obj)]) # 結果は文字列となることに注意
+ }
> ## シミュレーションによる例
> set.seed(123)
> x <- rpois(1000, lambda = 5) # 強度 5 の Poisson 乱数を 1000 個発生
> mode(x) # モードの計算
[1] "4"
> table(x) # 度数分布表の確認
x
  0   1   2   3   4   5   6   7   8   9  10  11  12  14
  7  29  89 135 189 167 155 100  62  35  20   9   2   1
> ## モードは数値以外のデータ (質的データ) にも適用可能
> ## 東京都 2016 年の日別最多風向データ fuko.csv による例
> ## 気象庁のホームページより取得
> ## http://www.data.jma.go.jp/gmd/risk/obsdl/index.php
> ## 東京都の 2016 年の各日の最多風向 (16 方位) を記録
> fuko <- read.csv("fuko.csv", fileEncoding = "sjis")
> head(fuko)
  月 日 最多風向
1  1  1   北北西
2  1  2     北西
3  1  3   北北西
4  1  4   西北西
5  1  5   北北西
6  1  6   北北東
> mode(fuko$最多風向) # モードの計算
[1] "北北西"
> table(fuko$最多風向) # 度数分布表の確認
西南西 西北西       東 東南東 東北東       南    南東 南南西 南南東       北    北西
     1     10      3      1     18     52     15     18     51     12     48
  北東 北北西 北北東
    27     85     25
```

(mode.r)

7.4. 参考文献

1. 東京大学教養学部統計学教室編「統計学入門」, 東京大学出版会 (1991 年).
2. 吉田朋広著「数理統計学」, 朝倉書店 (2006 年).