

Rで学ぶ確率統計

第3回 記述統計と確率

1

Rで学ぶ確率統計

第3回
記述統計と確率

講師 秋光 淳生



2

第3回 記述統計と確率

- 1 記述統計と代表値
- 2 確率
- 3 条件つき確率とベイズの定理



3

①-1 記述統計と推測統計

記述統計

- データから集団の性質を記述し要約する
- 代表値グラフ、表

推測統計

- データの奥の母集団の特徴を推測する
- 推定・検定・実験計画

尺度

変数

- 調べる対象項目
- 質的変数と量的変数

尺度

- 名義尺度
- 順序尺度
- 間隔尺度
- 比例尺度

4

①-2 名義尺度と順序尺度

名義尺度と順序尺度

名義尺度

他と区別するために用いる
(例) 学生番号、性別

順序尺度

値の大小関係がわかっている
(例) 好み、成績

成績一覧

1位	A	100点	50点
2位	B	50点	10点
3位	C	40点	30点
4位	D	10点	

コンピュータを操作することは

どちらかというと嫌い	嫌い	普通	好き	どちらかというと好き
1	2	3	4	5

Rで学ぶ確率統計

第3回 記述統計と確率

5

①-3 間隔尺度と比例尺度

間隔尺度と比例尺度

比が計算できるのは
0の位置が決まっているから

- 0に意味がある
(例) 距離 (長さが無い)
- 0に意味がない
(例) 時刻0時、摂氏0度

6

①-4 集団の特徴を表現する量(1)

分布の特徴を表現する主な値

- 平均値・分散と標準偏差
- 中央値・最頻値
- 5数要約
- 箱ひげ図
- Z値と外れ値
- 相関係数

- N 個の観測値 (量的変数)
 $x_1, x_2, \dots, x_N$
- この集団を表す特徴を知りたい
- 3つの代表値
平均と中央値と最頻値
- 最頻値 (モード)
度数分布表を作成した際に最も
度数の大きな階級の階級値
- 階級値
 a 以上 b 未満のときは $\frac{a+b}{2}$

7

①-4 集団の特徴を表現する量(2)

平均と分散

平均 (mean)

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

分散 (variance)

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

標準偏差 (Standard Deviation)

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Z値と外れ値

偏差 (平均との差) ($x_i - \bar{x}$)

$$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) = 0$$

Z値 (または Zスコア)

$$z_i = \frac{x_i - \bar{x}}{s}$$

Z値の平均0、分散が1 (標準化)

外れ値

他の観測値に比べ離れているもの

8

①-4 集団の特徴を表現する量(3)

5数要約

四分位数

観測値を大小で4分割した境界値

5数要約

- 最小値
- 第1四分位数
- 中央値
- 第3四分位数
- 最大値

箱ひげ図

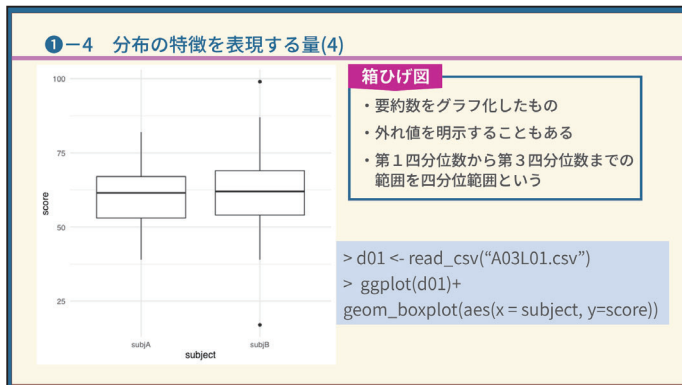
- 要約数をグラフ化したもの
- 外れ値を明示することもある
- 第1四分位数から第3四分位数までの範囲を四分位範囲という

```
> d01 <- read_csv("A03L01.csv")
> ggplot(d01)+
  geom_boxplot(aes(x=subject, y=score))
```

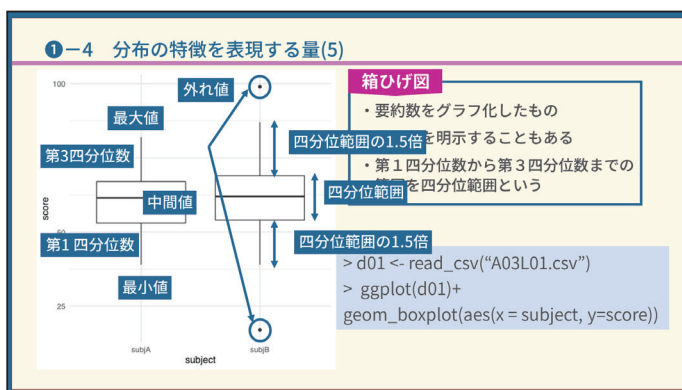
Rで学ぶ確率統計

第3回 記述統計と確率

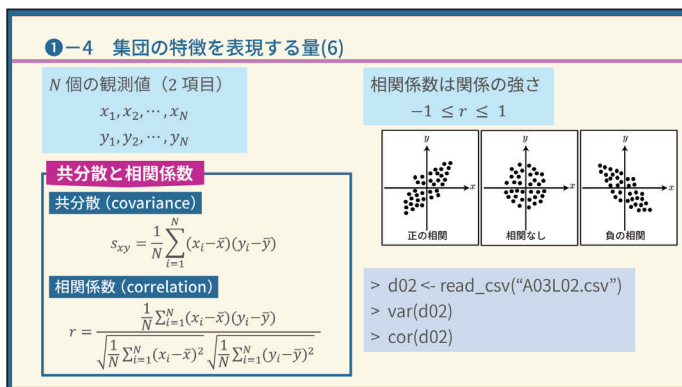
9



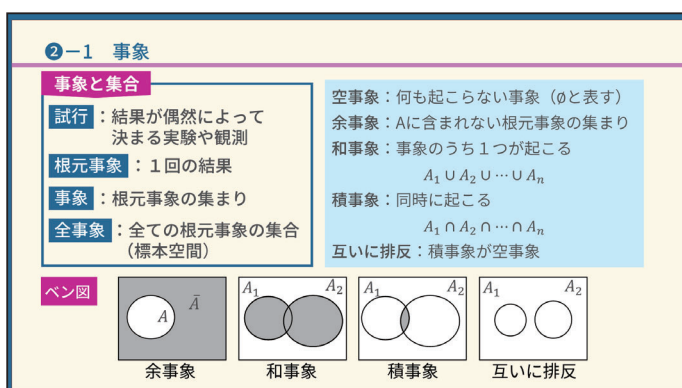
10



11



12



R で学ぶ確率統計

第 3 回 記述統計と確率

13

2-2 確率

確率

確率 偶然に起こる事象の起こりやすさ (定義は複数ある)

コルモゴロフの公理

事象 A について次の 3 つの性質を満たす 関数 $P(A)$ を確率という

1. 任意の事象 A について

$$0 \leq P(A) \leq 1$$

2. 全事象 S について

$$P(S) = 1$$

3. 事象がそれぞれ互いに排反のとき

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

一般には

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

14

2-3 確率の例

{1,2,3,4,5,6} の 6 個の目が等しい確率で
起こるサイコロを振ったときに
偶数が出る確率を求めよ

ある事象が起こるとき
変数 X の取りうる値の確率が
与えられているとき **確率変数** という
 $P(X = 1) = \frac{1}{6}$ と書く

偶数が出る事象を A とすると

$$A = \{2, 4, 6\}$$

どれも起こる確率は $\frac{1}{6}$ で、これらは同時に起こらない
(互いに排反なので)

$$P(A) = P(X = 2) + P(X = 4) + P(X = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

15

3-1 条件付き確率

条件付き確率と独立

事象 A が起こったという条件のもとで
事象 B が起こる確率を条件付き確率という
これを $P(B|A)$ とすると

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

確率の乗法定理 (上の式から)

$$P(A \cap B) = P(B|A)P(A)$$

が成り立つ

独立性

$$P(A \cap B) = P(A)P(B)$$

のとき、事象 A と B は独立であるという

独立

事象 A が起こったことが事象 B
の起こる確率に影響しない

16

3-2 ベイズの定理(1)

ベイズの定理

事象 A が起こる要因として互いに排反な事象 H_i ($i = 1, 2, \dots, n$) があり
 $P(A|H_i)$ がそれぞれ分かっているものとする

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}$$

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^n P(A|H_j)P(H_j)}$$

Rで学ぶ確率統計

第3回 記述統計と確率

17

③-2 ベイズの定理(2)

ベイズの定理

条件付き確率から

$$P(H_i|A) = \frac{P(H_i \cap A)}{P(A)}$$

$$P(A|H_i) = \frac{P(H_i \cap A)}{P(H_i)}$$

の2式から $P(H_i \cap A)$ を消す

また $P(A) = P(H_1 \cap A) + P(H_2 \cap A) + \dots + P(H_n \cap A)$
 $= P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + \dots + P(A|H_n)P(H_n)$
 とすると

$$P(H_j|A) = \frac{P(H_j)P(A|H_j)}{\sum_{j=1}^n P(A|H_j)P(H_j)}$$

18

③-2 ベイズの定理(3)

ベイズの定理

事象 A が起こる要因として互いに排反な事象 H_i ($i = 1, 2, \dots, n$) があり
 $P(A|H_i)$ がそれぞれ分かっているものとする

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}$$

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^n P(A|H_j)P(H_j)}$$

$P(H_i)$ を事前確率、 $P(H_i|A)$ を事後確率という
 (事象 A が起きたあとに原因が H_i であった確率を求める)

19

③-3 ベイズの定理の例題(1)

2つの箱があり、①の箱には4個のあたりくじ、5個のはずれくじが入っている。

②の箱には4個のあたりくじ、3個のはずれくじが入っている。

箱①、②のどちらかを選び、そこからくじを1つ取り出す。

箱を選ぶ確率を $1/2$ とする。箱の中からくじを取り出すかはどれも等しいとする。

取り出したくじがあたりくじであるとき、箱①から取り出された確率を求めよ。



20

③-3 ベイズの定理の例題(2)

H_1 : 箱①を選ぶ

H_2 : 箱②を選ぶ

A : あたりくじを取り出す

とすると

$$P(H_1) = P(H_2) = \frac{1}{2}, \quad P(A|H_1) = \frac{4}{9}, \quad P(A|H_2) = \frac{4}{7}$$

なので

$$P(H_1|A) = \frac{\frac{1}{2} \times \frac{4}{9}}{\frac{1}{2} \times \frac{4}{9} + \frac{1}{2} \times \frac{4}{7}} = \frac{7}{16}$$

