

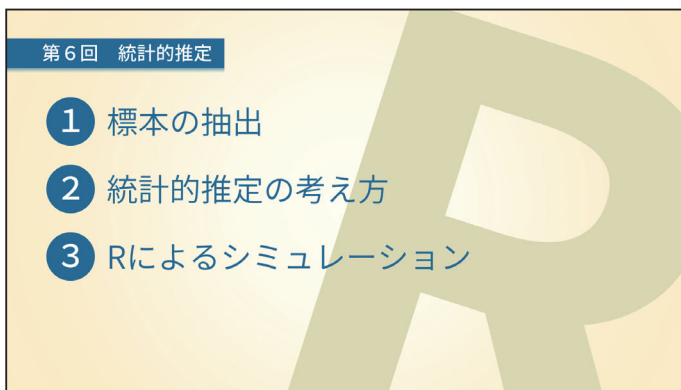
Rで学ぶ確率統計

第6回 統計的推定

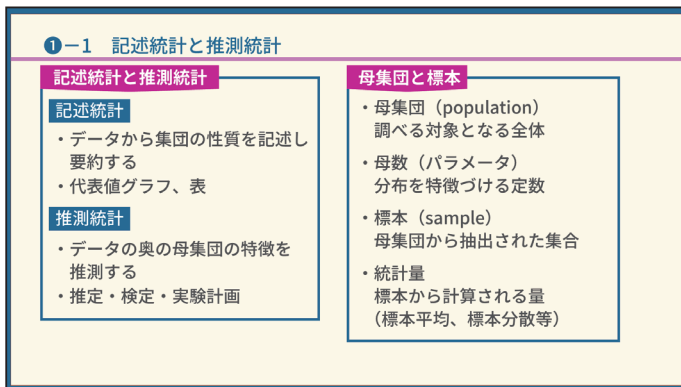
1



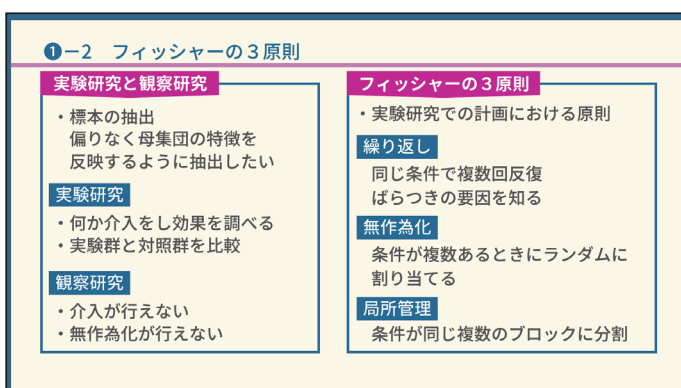
2



3



4



Rで学ぶ確率統計

第6回 統計的推定

5

①-3 標本の抽出

標本の抽出方法

単純無作為抽出法

全体から無作為に抽出

系統抽出法

1. 全体に番号をつける
2. 最初の人は無作為に選ぶ
3. その後等間隔で抽出

層化抽出法

性別や地域といった層ごとに抽出
層によるばらつきを減らす

多段抽出法

1. 多くのグループからいくつかを抽出
2. グループを構成する多くの小グループからいくつかの小グループを抽出
3. その小グループから何人かを抽出
といった段階を踏む

クラスター抽出法

特徴ごとに少数の集団（クラスター）に分けて、いくつかのクラスターを抽出
選ばれたクラスターには全数調査

6

②-1 点推定 (Point Estimation)

(例) 右のようなベルヌイ試行を n 回行う。各試行の結果を

$$X_i = \begin{cases} 1 & i \text{ 回目表} \\ 0 & i \text{ 回目裏} \end{cases}$$

とする

$$X = \sum_{i=1}^n X_i$$

とすると、 X は二項分布 $p(x) = {}_n C_x p^x (1-p)^{n-x}$ に従うと考えられる。母数 p を推定するために

$$\hat{p} = \frac{X}{n}$$

を用いるとする



7

②-2 点推定と区間推定

点推定と区間推定

- ・点推定（1つの値で推定）
推定量（観測値の関数）
推定値（実際の観測の結果）
- ・区間推定（区間を推定）

統計量

- ・統計量
標本から計算される量
- ・順序統計量
観測値を大きい順に並べ替えたもの
- ・十分統計量
パラメータを推定するのに十分な情報を含んだ統計量

8

②-3 推定量の判断基準

主な判断基準

- ・一貫性 (consistency)
観測数 n を無限にしたときに推定値がパラメータに確率収束

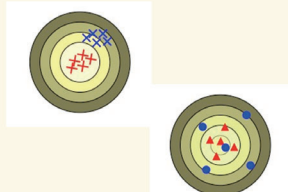
$$\hat{\theta} \xrightarrow{P} \theta$$

- ・不偏性 (unbiasedness)
推定量の期待値がパラメータと一致

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

- ・最小分散性 (minimum variance)
推定量の分散を最小にするもの

$$V[\hat{\theta}] = E[(\hat{\theta} - \theta)^2]$$



Rで学ぶ確率統計

第6回 統計的推定

9

2-4 (復習) 和の分布

和の分布

2つの確率変数 X, Y の和 $X + Y$ について・期待値 $E(X + Y) = E(X) + E(Y)$

・分散

$$\begin{aligned} V(X + Y) &= E[(x + y) - (\mu_x + \mu_y)]^2 \\ &= E[(x - \mu_x)^2] + E[(y - \mu_y)^2] + 2E[(x - \mu_x)(y - \mu_y)] \\ &= V[X] + V[Y] + 2Cov[X, Y] \end{aligned}$$

共分散

・互いに独立な場合
確率(密度)関数 $f(x, y) = f_x(x)f_y(y)$

$$\begin{aligned} P(X + Y = t) &= \sum_k f_x(t - k)f_y(k) \\ f(t) &= \int_{-\infty}^{\infty} f_x(t - y)f_y(y)dy = \int_{-\infty}^{\infty} f_x(x)f_y(t - x)dx \end{aligned}$$

10

2-5 標本平均と標本分散の期待値と分散(1)

標本平均の期待値と分散

・ n 個の確率変数 $X_i (i = 1, 2, 3, \dots, n)$ が独立に同じ分布に従う・ その分布の平均を μ 、分散を σ^2 とする・ 標本平均を $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ とすると

・ 期待値

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} (\mu + \mu + \dots + \mu) = \mu$$

・ 分散

$$V(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} (\sigma^2 + \sigma^2 + \dots + \sigma^2) = \frac{1}{n} \sigma^2$$

また

$$\begin{aligned} E(\sum_{i=1}^n X_i^2) &= E(\sum_{i=1}^n (X_i - \mu + \mu)^2) \\ &= E(\sum_{i=1}^n (X_i - \mu)^2 + 2\mu \sum_{i=1}^n (X_i - \mu) + \sum_{i=1}^n \mu^2) = n(\mu^2 + \sigma^2) \end{aligned}$$

11

2-5 標本平均と標本分散の期待値と分散(2)

標本平均の期待値と分散

標本平均の分散

$$\begin{aligned} E(\bar{X}^2) &= E((\bar{X} - \mu + \mu)^2) = E((\bar{X} - \mu)^2 + 2\mu(\bar{X} - \mu) + \mu^2) \\ &= E((\bar{X} - \mu)^2) + E(2\mu(\bar{X} - \mu)) + E(\mu^2) \\ &= E\left(\sum_{i=1}^n \frac{1}{n^2} (X_i - \mu)^2\right) + 0 + \mu^2 = \frac{1}{n} \sigma^2 + \mu^2 \end{aligned}$$

ここから

$$E(\sum_{i=1}^n (X_i - \bar{X})^2) = E(\sum_{i=1}^n X_i^2) - nE(\bar{X}^2) = (n-1)\sigma^2$$

 $\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 不偏分散

12

2-6 区間推定

区間推定

・ 点推定は推定値を1つ求める

・ 区間推定では区間を求める

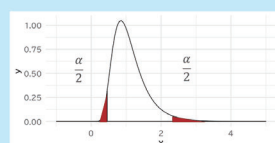
・ 信頼係数

真の値を含む確率
95% や 99% など

・ 信頼区間

実現値から求められた区間

- ・ 未知の母数を含む母集団の分布
- ・ 推定量の分布から信頼区間が求まる
- ・ 実際の母数が $[\theta_L, \theta_U]$ の区間に
100(1 - α)% の確率で含まれる



第6回 統計的推定

17

③-1 区間推定(2)

#後半 (1)のつづき

z<- tibble(count=1:m,est= est,

lower=est - qnorm(1-alpha/2,0,1)* sqrt(est*(1-est)/n),

upper=est + qnorm(1-alpha/2,0,1)* sqrt(est*(1-est)/n),

check= ifelse(abs (prob-est) < qnorm (1-alpha/2 ,0,1) *
sqrt(est*(1-est)/n) , "Y", "N")

#qnorm(0.975,0.1) で
累積確率が 0.975

ifelse (条件,真,偽)

#グラフ

ggplot(z)+geom_point(aes(x=count,y=est))+

geom_segment(aes(x=count,y=lower,xend=count,yend=upper,colour=check))+

geom_hline(yintercept=prob,colour="red")

#geom_segment
(x,y) から(xend,yend)

5

10

15

20

25

30

35