

統計の基礎概念



確率の基本的事項

確率の基本的事項

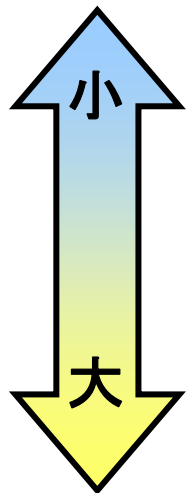
- 確率0 : ある事柄が起こる可能性が全くない場合
- 確率1 : その事柄が必ず起こる場合
- $0 \leq \text{確率} \leq 1$
- 事象Aが起こる確率を“ $\text{Pr}(A)$ ”と書く

例: 女子の生まれる確率を調べる

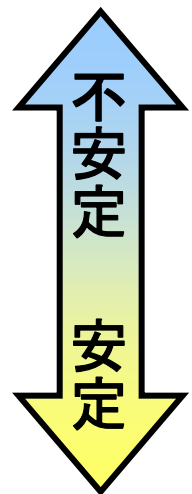
- 新生児100人を調査 \Rightarrow 100人中54人が女子
- 「女子の生まれる確率は0.54である」といえるか？
- 「確率」という言葉は、より多くの場合を調べてから用いるのが一般的

男児と女児の出生率とその割合

抽出する人数が多いほど、割合は安定してくる



抽出の人数	男児の数	女児の数	男児の割合	女児の割合
100	46	54	0.46	0.54
500	244	256	0.488	0.512
1000	533	467	0.533	0.467
2500	1288	1212	0.5152	0.4848
5000	2552	2448	0.5104	0.4896
10000	5118	4882	0.5118	0.4882



参考：1998年から2002年に生まれた女子の割合

$$2,860,580 / 5,876,979 = \underline{0.487}$$

確率

確率：割合から理想化された値

確率の表し方

- 事象A：女子が生まれる
 - ⇒ 事象Aがおこる確率 = 女子が生まれる確率
 - ⇒ $\Pr(A) = 0.487$
- 事象B：男子が生まれる
 - ⇒ 事象Bがおこる確率 = 男児が生まれる確率
 - ⇒ $\Pr(B) = 0.513$
- 全事象の確率の総和は1
 - ⇒ $\Pr(A) + \Pr(B) = 1$

確率変数と確率分布

最初に女子, 続いて男子が生まれる確率

- 考え得る事象：（第一子の性別, 第二子の性別）
 - (女, 女)－事象(A, A)
 - (女, 男)－事象(A, B)
 - (男, 女)－事象(B, A)
 - (男, 男)－事象(B, B)
 - 女子が生まれる確率：0.487
 - 男児が生まれる確率：0.513

最初に生まれた新生児の性別とは無関係(**独立**)に,
2番目に男子あるいは女子が上記の確率で生まれるとする.

確率変数

□ 確率変数

- 変数の値に確率に対応している

□ 女子の生まれた人数を x (確率変数) とする.

- $\Pr(x=0) = \Pr\{(\text{男}, \text{男})\}$

$$= \Pr(\text{男}) \times \Pr(\text{男}) = 0.513 \times 0.513 = 0.263$$

- $\Pr(x=1) = \Pr\{(\text{男}, \text{女})\} + \Pr\{(\text{女}, \text{男})\}$

$$= 0.513 \times 0.487 + 0.487 \times 0.513 = 0.500$$

- $\Pr(x=2) = \Pr\{(\text{女}, \text{女})\}$

$$= \Pr(\text{女}) \times \Pr(\text{女}) = 0.487 \times 0.487 = 0.237$$


確率分布

確率分布

- 確率変数 x の値と, x がそれぞれの値をとる確率との対応関係を示したもの
- 確率変数 x のとりうるすべての値 i について, 確率 $\Pr(x=i)$ の総和はつねに1
- 確率変数 x のとる各値に対して, 確率がどのように配分されるかという様子がわかる

x	確率
0	0.263
1	0.500
2	0.237
計	1.000

母集団と標本



標本調査の例

日本人の18才の女性の身長を調査

- 調査対象となる女性の総数：約56万人（母集団）
 - 母集団：調査あるいは研究対象の集まり

- 調査方法1
 - 約56万人の身長を調査：不可能に近い

- 調査方法2
 - 費用と時間の許す範囲内で、出来るかぎり多くの人を抽出し、その人達の身長を測定することによって、18才女性の身長（母集団）の特徴を推測する：**現実的な方法**

標本

□ 標本：観測値の集まり

□ 標本の例

- 18才の女性の中から100人を抽出し、身長を測定したところ次のようであったとする
161.5, 152.7, 163.5, ..., 161.7, 159.2, 163.4
- この100人について測定した数値から、母集団である18才の女性の身長の特徴を推測する

標本抽出の仕方

日本人の18才女性の身長の平均値を調べる

- 標本：日本全国から100人を抽出
 - 偶然、関東地域と近畿地域の人が多く抽出されたとする
 - 問題点：これらの地域の身長の平均が全国平均より大きいならば、結果として本当の値よりやや大きめの値を全国平均としてしまうおそれがある

母集団の特徴をよく把握している標本を抽出するにはどのようにしたら良いか？

偶然生じる偏りの回避

偶然に生じる偏りを少なくするための方法

- 日本のある地域の平均は全国平均より0.7cm低く、別の地域では0.6cm程高いことを知っているとする
- 各地域から、18才女性の人口数に比例して標本を抽出すると、標本抽出の偶然によって起こるこの種の偏りを少なくすることができる
- もっている知識を利用することによって偏りのない標本を抽出することができる

無作為抽出



患者の看護に対する意見調査

目的：ある療養所でより良い看護をしたい

□ アンケート調査を行い、意見調査を行う

- 病床数は1100
- 調査時間や集計費用の都合上、全ての患者を調べることはできない
- 一部の患者を調べて、入院患者全員(母集団)の意見を推測する

□ 調査項目

- あなたが食欲のないときに看護師は親切に聞いてくれますか
- あなたの検温や検脈のときに看護師は容態を聞いてくれますか
- あなたは平等な看護を受けていると思いますか

等

標本抽出の方法

□ 標本抽出の方法

- 療養所では5人が一部屋に入っている。
各部屋から2人ずつ選んで質問をした。
- 221号室には正夫, 明夫, 和夫, 信夫, 晴夫の
5人が入院している。
- この中から2人を選ぶとき, 選ぶ人の主観が入らない
客観的な抽出方法によって選ぶことが重要
- どのような方法で2人を選べばよいただろうか？

無作為抽出

□ 2人の選び方

(正夫, 明夫) (正夫, 和夫) (正夫, 信夫) (正夫, 晴夫)
(明夫, 和夫) (明夫, 信夫) (明夫, 晴夫)
(和夫, 信夫) (和夫, 晴夫)
(信夫, 晴夫)

□ 正夫が選ばれる確率 : $4/10=0.4$

- 明夫, 和夫, 信夫, 晴夫についても, それぞれが選ばれる確率は0.4

□ 無作為抽出

- 5人の中の誰もが選ばれるチャンスを等しくするような抽出方法
- くじによる抽出 : 当たる確率が0.4であるような, 5本の内2本が当たりくじになっているくじを作り, くじを引いて当たった人2人を抽出