

確率と統計

中山クラス 第5週

0

試験に関する注意

- ◆携帯電話はアラームを解除して、電源を切っておくこと。
呼び出し音が鳴るのは試験妨害
切るための操作は不正行為の疑い
- ◆イヤホン、MP3プレーヤやMDプレーヤ等は鞆にしまうこと。
- ◆荷物:まとめて床、教室右手の棚または後ろの席に置く
机に出して良いもの:筆記用具、時計(携帯は×)、学生証のみ。
- ◆問題用紙(1部)、答案用紙(2枚)と計算用紙(1枚)を配布。
- ◆「はじめ」の合図があるまで問題用紙は伏せておくこと。
「解答やめ」の合図があったらすぐに解答をやめること。
- ◆答案用紙を回収する。
最後尾の席の学生が前のほうに回収して教員に渡す。
答案用紙2枚を回収する。

3

今日の内容

- ◆小テストについて
- ◆予想問題の解説
- ◆第4章 母集団と標本
 - 4.1 母集団と標本
 - 4.2 推定統計の分類
 - 4.3 点推定
- ◆自習、コンピュータ演習、質問

1

不正行為について

不正行為と疑われそうなことはしない。

- ・ 後ろや左右を見ないこと。
- ・ 他の学生から見えるように答案用紙を置かないこと。
- ・ ポケットからティッシュやハンケチを出す場合は拳手して、教員の許可を得ること。
- ・ 鉛筆や消しゴムを落とした場合は自分で拾わず拳手すること。
- ・ 鞆の中の教科書やノートの内容が見えないこと。
- ・ 机の中に何も入れないこと。自分のものでないものが入っている場合は、一旦、自分の鞆にしまい、試験終了後に机の中に戻す。

6

小テストについて

11月7日(木)5限

試験時間:60分

試験範囲:教科書第1章～3章

持ち込み:筆記用具、時計、学生証のみ

試験の内容:予想問題を完全に解けるように!

2

不正行為の扱い

不正行為があった場合、例外なく(「初めてなので...」「他の科目ではやっていないので...」などと言われても受け付けない。)以下のように扱う。

1. その科目の小テスト、達成度確認試験およびレポート点はすべて0点とし、不合格とする。以上はこの大学では相応と考えられていて、これまでも実際そのように扱っています。
2. 教務課および修学相談室に連絡する。以後の処分は大学側に従うこと。

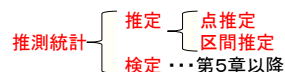
このようなことにならないよう、不正行為をしないこと。

7

予想問題の解説

8

4.2 推測統計の分類



推定: 具体的な値を用いて「母数は〇〇くらいだろう」
 日本の中学生全体の平均点は60点位だろう
 →点推定
 日本の中学生全体の平均は50点から70点くらい
 →区間推定

検定: 母集団について述べた異なる立場の2つの主張
 (仮説)のうち、どちらを採択するかを決めるもの。
 「平均的な学力は5年前から変化していない」 仮説A
 「平均的な学力は5年前から変化した」 仮説B

11

第4章 母集団と標本

3章まで: 手元にあるデータの様子を記述する方法

4章から: 大きな集団から一部を取り出した少数のデータの情報を使って、元の集団の性質について推測→**推測統計**

9

4.3 点推定

4.3.1 点推定の手順

1. 17才の日本人男性全体の平均身長を推定
2. 10人を標本として抽出
→ **サンプルサイズ**, **標本の大きさ**=10
3. 10人の身長データ(165.2~171.3cm)を用いて平均を計算
4. 平均=169.36cm
5. 17才の日本人男性全体の平均身長=169.36cmと推定。

12

4.1 母集団と標本

非常に大規模なデータ全体(→母集団)の
 (日本国民全体や工場で生産される製品全体に関するデータ)

統計的な性質(→母数)を
 (度数の比率, 平均, 分散, 相関など)

対象の一部を取り出したデータ(→標本)から推測

標本抽出: 母集団から標本を取り出すこと。

母数: 母集団の性質を表す統計的指標(比率, 平均, 分散, 相関係数など)

10

4.3.2 推定量と推定値

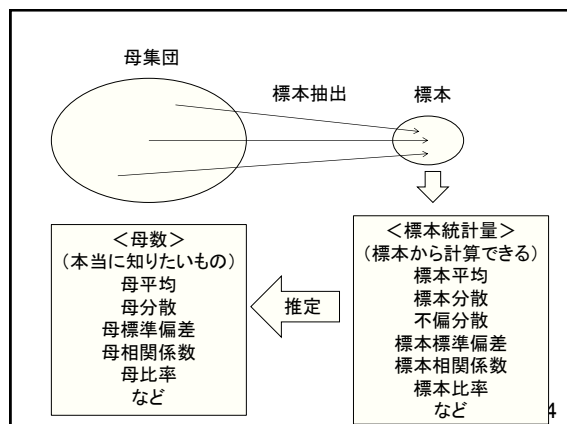
標本統計量: 標本データから計算される統計量
 標本平均, 標本分散, 標本相関係数など

母集団の統計量:
 母平均, 母標準偏差, 母相関係数など

母数の推定量:
 母数を推定するために用いられる標本統計量
 計算式(関数)や名前(平均など)

推定値: 標本データを用いて計算された推定量の値
 (数値)

13



4.3.3 標本抽出に伴う誤差

実際の母数の値にどの程度近い推定値が得られるか？
推定の結果はどのくらい信用できるか？

標本誤差: 抽出された標本データのみを用いることで生じる誤差

(例) 母集団=[1, 2, 3, 4, 5] 母平均=3
標本データ=[1, 4] 標本平均=2.5
標本データ=[2, 5] 標本平均=3.5

推定の結果「この程度の誤差が生じる」ことを知ることが重要。→ **標本分布** ... 4.4節以降で取り上げる

17

母数と推定量

母数	推定量	推定値
母平均	標本平均	標本データから計算
母分散	不偏分散	標本データから計算
母標準偏差	不偏分散の正の平方根	標本データから計算
母相関係数	標本相関係数	標本データから計算
母比率	標本比率	標本データから計算

15

不偏分散について

nサンプルの標本データ

$$X = [x_1, x_2, \dots, x_n]$$

標本平均

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

標本分散

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

nサンプルの標本データを多数回抽出(独立試行)する

→ x_i に対して多くのデータを得る

→ 集合平均(期待値) $E[x_i]$ を計算 → 母集団の平均 (μ)

同様に, $E[(x_i - \mu)^2]$ を計算 → 母集団の分散 (σ^2)

3

母比率／標本比率

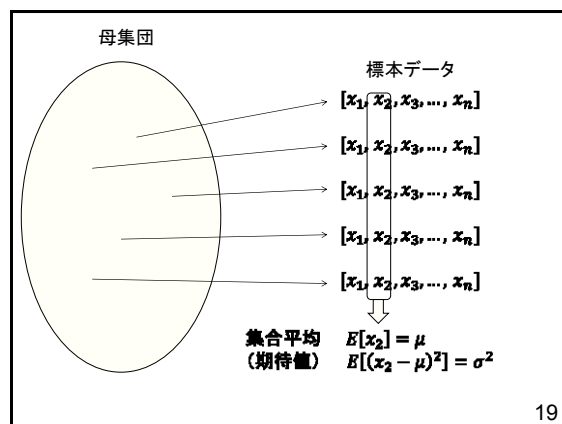
比率: ある特徴を持った人や物が全体の中に含まれる割合

(例)

ある学校の生徒にしめる男子生徒の割合
日本全体の有権者全体の中で現内閣を支持する人の割合
あるテレビ番組を見ている世帯の割合(視聴率)

内閣支持率やテレビの視聴率などは、標本抽出に基づいた母数の推定が行われている。

16



19

標本平均の期待値, μ は母集団の平均

$$\begin{aligned} E[\bar{x}] &= E\left[\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right] \\ &= \frac{1}{n}(E[x_1] + E[x_2] + \cdots + E[x_n]) \\ &= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu \end{aligned}$$

標本平均の期待値=母集団の平均

標本平均の散らばり, σ^2 は母集団の分散

$$\begin{aligned} E[(\bar{x} - \mu)^2] &= E\left[\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n - n\mu)\right)^2\right] \\ &= E\left[\left\{\frac{1}{n}((x_1 - \mu) + (x_2 - \mu) + \cdots + (x_n - \mu))\right\}^2\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n E[(x_i - \mu)^2] = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{1}{n}\sigma^2 \end{aligned}$$

△U

標本分散の変形

$$\begin{aligned} s^2 &= \frac{1}{n} \sum_{i=1}^n \{(x_i - \mu) - (\bar{x} - \mu)\}^2 \\ &= \frac{1}{n} \sum_{i=1}^n \{(x_i - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2\} \end{aligned}$$

第2項の変形

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

21

標本分散の期待値

$$\begin{aligned} E[s^2] &= \frac{1}{n} \sum_{i=1}^n E[(x_i - \bar{x})^2] - E[(\bar{x} - \mu)^2] \\ &= \sigma^2 - E[(\bar{x} - \mu)^2] = \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2 \end{aligned}$$

母集団の分散 σ^2 と標本分散の期待値 $E[s^2]$ の関係

$$\sigma^2 = \frac{n}{(n-1)} E[s^2]$$

母集団の分散は不偏分散となる。

22