統計学2及び演習

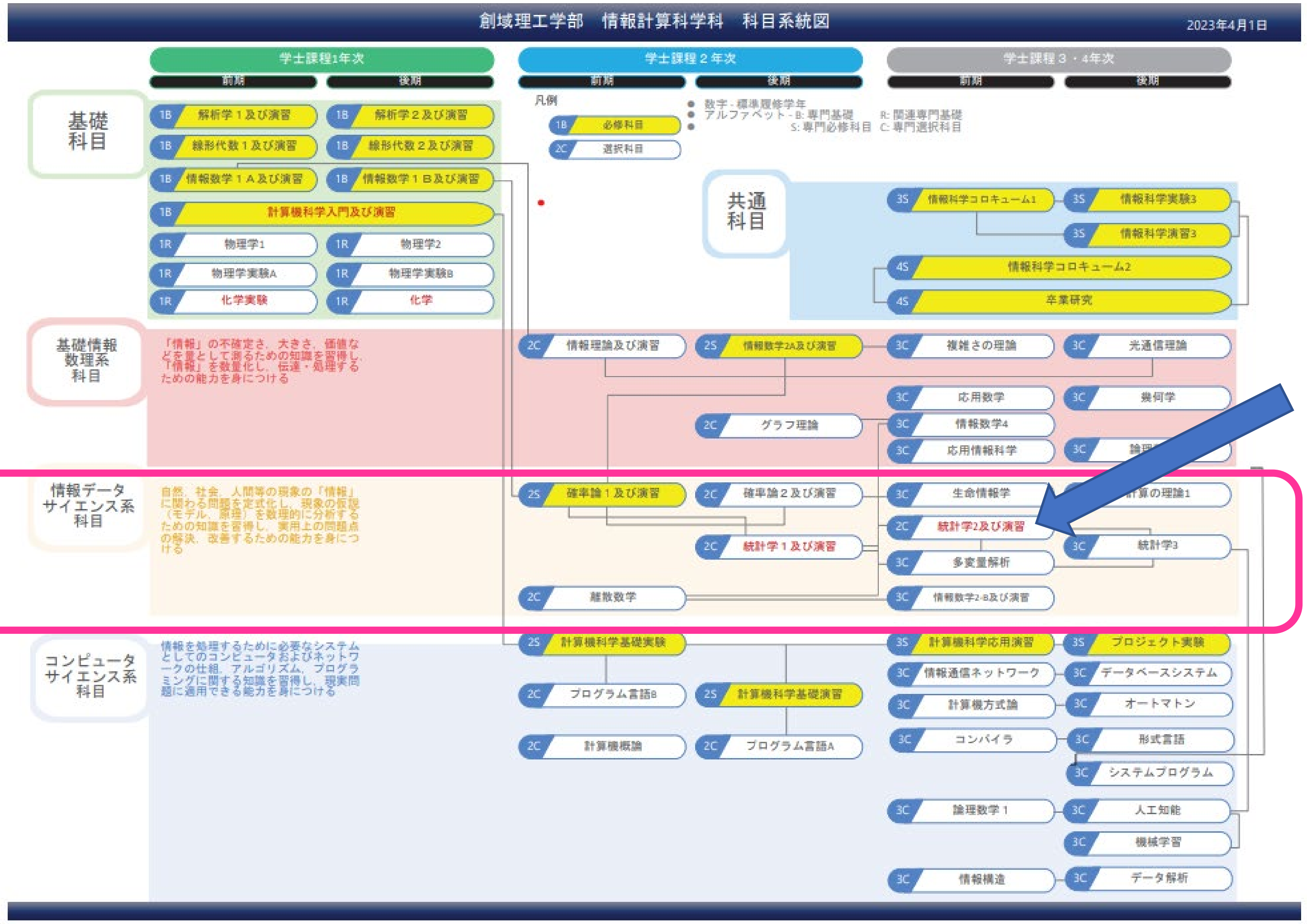
ガイダンス

創域理工学部
Faculty of Science and Technology

東京理科大学
創域理工学部情報計算科学科
安藤宗司

2023年4月12日

情報計算科学科の科目系統図



統計学2はここ！

情報データサイエンス系科目に注目すると

- 統計学1及び演習（2年後期）
  - 統計的推測の推定を扱った
- 統計学2及び演習（3年前期）
  - 統計的推測の検定を扱う
- 多変量解析（3年前期）
  - 互いに従属する一群の変量間の関係を扱う手法を扱う
- 統計学3（3年後期）
  - 線形・非線形回帰モデル
  - 判別分析，主成分分析，クラスター分析
- データ解析（3年後期）
  - 一般化線形モデル
  - 生存時間解析
  - サンプルサイズ設計

統計学2のカリキュラム

	日程	内容
第1回	4/12	ガイダンス 検定の考え方（帰無仮説，対立仮説，棄却域，有意水準）
第2回	4/19	第1種の誤り，第2種の誤り，検出力
第3回	4/26	最強力検定，ネイマン・ピアソンの補題とその例
第4回	5/10	一様最強力検定とその例
第5回	5/17	不偏検定，相似検定とその例
第6回	5/24	母比率の検定
第7回	5/31	尤度比検定とその例
第8回	6/7	適合度検定とその例

統計学2のカリキュラム

	日程	内容
第9回	6/14	適合度検定とその例
第10回	6/21	分割表における独立性検定とその応用例
第11回	6/28	線形回帰分析の考え方
第12回	7/5	最小二乗法とその性質
第13回	7/12	一般化最小二乗法，最尤法とその性質
第14回	7/19	多重共線性，Ridge回帰
第15回	7/26 or 8/2	到達度評価試験

成績評価方法

- 到達度評価試験（60%）
- 中間試験（40%）

統計学2及び演習

検定の考え方

創域理工学部
Faculty of Science and Technology

東京理科大学
創域理工学部情報計算科学科
安藤宗司

2023年4月12日

Contents

- 検定の考え方
- 仮説の設定
  - 帰無仮説，対立仮説
- 棄却域
- 有意水準

手元にあるコインはいかさまコインかどうか

- 表が出る確率 π は1/2かどうか
- 実際にコインをN 回投げて，確かめる実験を考える

表 表 表 表 表 表 … 表 n 回

裏 裏 裏 裏 裏 裏 … 裏 $N-n$ 回

- 表が出た割合 $p=\dfrac{n}{N}$

コイン投げの実験を仮説検定で検証

- 仮説の設定
  - 表が出る確率は1/2かどうか

- 設定した仮説を評価するためのデータを収集
  - 実際にコインをN回投げる

    ↳ Nを設定することをサンプルサイズ設計という

- 事前に設定した判定基準に基づき判断
  - 表がm回以下（または以上）のとき，いかさまコインと判断

    ↳ この判定基準を確率論的に設定する

仮説の設定

- 「表が出る確率 π は1/2かどうか」を検証するために2つの仮説を設定する
  - 帰無仮説 $H_0 : \pi = 1/2$
  - 対立仮説 $H_1 : \pi \neq 1/2$

- 帰無仮説が成り立つと仮定する
  - 手元にあるコインはいかさまコインではないと仮定する
  - 収集したデータに基づき帰無仮説が成り立つかどうかを判断する

設定した仮説を評価するためのデータを収集

- 「帰無仮説」と「対立仮説」のどちらが正しいかを判断するために，コインを N 回投げる
  - N（サンプルサイズ）はどう設定すればいいのだろうか？
- 検出力に基づいてサンプルサイズを設計する
  - 検出力は検定の精度を表す指標
  - 詳しくは後ほど紹介

事前に設定した判定基準に基づき判断

- 判断基準の考え方
  - いかさまコインではないとき
    10回コインを投げれば常に表が5回出るとは限らない
  - 表が出る回数は確率的に変動する
  - 偶然に出る可能性のある「表の回数」の範囲を考える

- この範囲を確率論的に設定する

偶然に出る可能性のある「表の回数」の範囲

- いかさまコインではない（帰無仮説 $H_0: \pi = 1/2$）と仮定

表の回数	確率
0	0.1%
1	0.98%
2	4.39%
3	11.72%
4	20.51%
5	24.51%
6	20.51%
7	11.72%
8	4.39%
9	0.98%
10	0.1%

確率の計算式 $\quad {}_{10}C_x \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{10-x}$

表の回数3から7である確率は85%以上

表の回数が1以下，または9回以上である確率は2.16%
表の回数が2以下，または8回以上である確率は10.94%

有意水準

- 帰無仮説のもとで，5%（または1%）未満でしか起きない事象は偶然ではないと考える

- 10回コインを投げた結果
  - 表の回数が1以下，または9回以上の場合 ➡ 偶然ではない
  - 表の回数が2以上，または8回以下の場合 ➡ 偶然である

- 棄却域
  - 偶然ではないと考える範囲

検定結果の解釈

- 10回コインを投げた結果
  - 表の回数が1以下，または9回以上の場合
    - 統計学的に有意と判定
    - 帰無仮説を棄却して，対立仮説を採択する
    - 「表が出る確率 π は1/2 ではない」と判断する
  - 表の回数が2以上，または8回以下の場合
    - 統計学的に有意でないと判定
    - 帰無仮説を採択する
    - 「表が出る確率 π は1/2 ではない」とはいえないと判断する

「表が出る確率 π は1/2 である」とは判断できないことに注意！

統計的仮説検定の一般論

- 記号の定義
  - 母集団分布Pからの無作為標本 $X_1, X_2, \ldots, X_n$
  - 観測結果として生じうる全体の集合（標本空間） $\mathcal{X}\ (\subset \mathbb{R}^n)$
  - 実際の観測結果 $(x_1, x_2, \ldots, x_n)$
    集合$\mathcal{X}$の値をとる標本ベクトル$(X_1, X_2, \ldots, X_n)$の1つの実現値
  - 母数空間 $\Theta\ (\subset \mathbb{R}^s)$
  - パラメータベクトル $\theta = (\theta_1, \theta_2, \ldots, \theta_n) \in \Theta$

仮説の設定

- パラメータ空間の分割

$$\Theta = \Theta_0 \cup \Theta_0^c \text{ かつ } \Theta_0 \cap \Theta_0^c = \phi$$

- 帰無仮説 (null hypothesis)

$\theta \in \Theta_0$という命題

- 対立仮説 (alternative hypothesis)

Θ_0^c の部分集合Θ_1に対して，$\theta \in \Theta_0$という命題

仮説の選択問題

- 仮説

帰無仮説 $H_0: \theta \in \Theta_0$　　対立仮説 $H_1: \theta \in \Theta_1$

- 問題
  - どちらかの仮説が正しいと仮定
  - 帰無仮説と対立仮説のどちらかを決定する問題を考える

- 仮説の種類
  - Θ_0が1点のとき，単純仮説 (simple hypothesis)　$\Theta_0 = \{\theta_0\}$
  - Θ_0が2点以上のとき，複合仮説 (composite hypothesis)
  - Θ_1に関しても同様　$\Theta_0 = \{\theta \in \Theta \mid \theta \leq \theta_0\}$

仮説の棄却

- 標本空間の分割
  - 標本空間$\mathcal{X}$の部分集合W　　$\mathcal{X} = W \cup W^c$　$W \cap W^c = \phi$

- 帰無仮説の棄却

  $(x_1, x_2, \dots, x_n) \in W$ならば帰無仮説を棄却する

- 帰無仮説の採択

  $(x_1, x_2, \dots, x_n) \in W^c$ならば帰無仮説を採択する

- 棄却域をWと採択域をW^cという