

データの分布

樋口さぶろお

龍谷大学工学部数理情報学科

確率統計☆演習 I L01(2017-09-20 Wed)

最終更新: Time-stamp: "2017-09-20 Wed 07:38 JST hig"

今日の目標

- e ラーニングシステムで学習できる
- メールで連絡を受け取れる
- データから, 手で度数分布表とヒストグラムが作れる



<http://hig3.net>

ここまで来たよ

1 はじめに

- この授業どんなのり?

2 データの分布

- データとは?
- 度数分布表
- ヒストグラム

学習目標

講義概要 → シラバス

現実世界の現象を理解し、数理モデルとの関係を明らかにするためには、観察・実験により取得したデータを整理・解析することが必要です。データを整理して表現する記述統計と、限られたデータから数理モデルのパラメタを推測する推測統計を説明します。ただし、量的1変数の場合を主に扱います。これに必要な範囲で確率論を説明します。数式を用いた解析、ソフトウェアによる解析の両方に習熟してもらいます。

到達目標 → シラバス

- 実験・観察により取得した量的1,2変数データを統計的に整理して表現し、他者に対して説明できる。
- データから数理モデルのパラメタを推測して、根拠とともに他者に説明できる。
- データから仮説を立てて検証し、他者を説得できる。

確率統計☆演習 I を履修してはいけない理由

次のどれも響かない人は履修しないことを奨めます。

- コア選択必修 M
- (3 年前期) 確率統計☆演習 II, 計算科学☆実習 B の前提科目
- 数学の教員免許の必修科目
- 新課程の 高校の **高校 数学 I (データの分析)**=毎年センター試験に出題, **高校 数学 A (場合の数と確率)**, **高校 数学 B (確率分布と統計的推測)**
- 教育の評価に統計は必要
- いま, 統計学が熱い!
- いま, ビッグデータ, 人工知能 (AI), 機械学習 (machine learning) が熱い!!
- 統計は科学技術の言葉 ⇨ 数理卒は当然期待されてる
- 統計検定 2,3 級

こんなことに答えます

- ① 高校の数学で、こういう教え方導入したら、ちょっとだけ平均点が上がった。これ効果あったって言うていいの?
- ② YouTube から猫の動画を見つけるアルゴリズム、こう改良して、100個の入力画像で試したら、判定精度がちょっとあがった。これで結論だしていいの? 10000個でやり直すべき?n
- ③ 秋元PはチームAにチームKより身長高いメンバーをいれてる説を唱えたけどみんな信じてくれない…どうやって説得する?

確率統計☆演習 I ののり

成績計算難しくないけどとにかく注文の多い科目です…

科目の成績 100 ピーナッツは

- 25 ピーナッツ: 毎回授業での非参照 quiz, e ラーニングの予習問題, 授業時間内の活動, それほどたいへんじゃないレポートなど
- 30 ピーナッツ: プチテスト (11 月) <http://www.toukei-kentei.jp> 3 級で一部分を代用可能.
- 45 ピーナッツ: ファイナルトライアル (定期試験期間)
- その他追加ピーナッツ. その時に説明.

その時点のピーナッツにかかわらず, ファイナルトライアルに参加しないと合格にはなりません. ファイナルトライアル時点で 15 ピーナッツ未満の人も, (平均点を上げるために) 参加をすすめますが, 追試験はなし.

欠席届 毎回出席を前提に進めます. やむを得ず欠席して, ピーナッツ的に考慮されたい場合は, 専用用紙に事情を説明する書類を貼って, 授業前後各 5 分に提出 (事前事後とも可. ファイナルトライアルが締切). 欠席に事前連絡は原則不要. 何回欠席してもファイナルトライアル参加資格を失うことはありません.

担当者ののり

- なまえ: 樋口さぶろお `hig-probstat@math.ryukoku.ac.jp`
- へや: 1-502
- 樋口オフィスアワー月 3.5(1-539) 金 4(1-502), Math ラウンジ月-木昼 (1-614)
- Web ページ: <http://hig3.net> (表紙に QR コード) 演習の指示や, スケジュールもここから.

教科書やその他の準備

必須です. 西川確率統計

<https://www.amazon.co.jp/product/dp/4781913555>

ソフトウェア表計算ソフトウェア Microsoft Excel を演習で使います. 龍大生は Office365 で無料で入手可能. <http://portal.office.com>
電卓平方根 (ルート) のあるもの. スマホアプリでも可.

1 週間のタイムライン

- ① 木 09:00 ごろまでに 予習復習問題 (=Trial 予想問題) を Learn Math Moodle で公開 Trial までの最高点を記録.
- ② 水 1 の最初 Trial(=小テスト) 参照不可 相談不可 (水 1 までにやったことについて, 水 1 の最後に Trial するパターンも)
- ③ 水 1 チーム別エリア座席指定. 講義のような演習のような. チームで何かやったり.
- ④ 水 1 の最後 来週の Trial の予告
- ⑤ 樋口オフィスアワー月 3.5(1-539) 金 4(1-502), Math ラウンジ月-木昼 (1-614)

Learn Math Moodle を使ってみよう

<http://hig3.net> → Learn Math Moodle (全学認証) → 確率統計☆演習 I

<http://hig3.net> → 確率統計☆演習 I → 配布資料.

メール通知多数 スマホで <https://g.mail.ryukoku.ac.jp> Gmail アプリ

ここまで来たよ

- 1 はじめに
 - この授業どんなのり?
- 2 データの分布
 - データとは?
 - 度数分布表
 - ヒストグラム

1 変数の量的データ

某アイドル集団群+某バレーボール選手 (1名) の身長データ.

148cm
148.5cm
149cm
⋮
185cm

<http://girlsgroupguide.com/>

このコースの最後までいくと問えること (正確な表現ではありません)

- オーディションにおいて、身長は考慮されているか?
- チーム編成において、身長は考慮されているか?
- ⋮

ここまで来たよ

- 1 はじめに
 - この授業どんなのり?
- 2 データの分布
 - データとは?
 - 度数分布表
 - ヒストグラム

度数分布表

高校 数学 I 西川確率統計 §5.1.1

階級	度数	相対度数
145 より大きく 150 以下	7	0.09
150 より大きく 155 以下	17	0.22
155 より大きく 160 以下	29	0.37
160 より大きく 165 以下	19	0.24
165 より大きく 170 以下	4	0.05
170 より大きく 175 以下	1	0.01
175 より大きく 180 以下	0	0.00
180 より大きく 185 以下	1	0.01
185 より大きく 190 以下	0	0.00
合計	78	1.00

- 階級幅は一定で
- 相対ナントカ (比率) の合計が 1 にならないとき. 度数分布表と限らず一般に, 無視して 1 と書くか, 相対誤差が小さい行で調整.
- 以下, 以上, 未満 (=より小さい), より大きい

度数分布表の作り方

高校 数学 I 西川確率統計 §5.1.1

- **階級** = 一定間隔で区切った区間, 下品な?言葉 'bin' ビン. いくつに分けるか? 一概には言えないけど, 切りのいい値にしちゃっていい.
- **階級幅** = 区間の幅
- **階級値** = その階級のまん中の値
- **度数** = その範囲に入ってるデータの個数
- データ全体の個数 = 度数の合計 = n
- **相対度数** = 度数 / データ全体の個数 (%で書くことも)

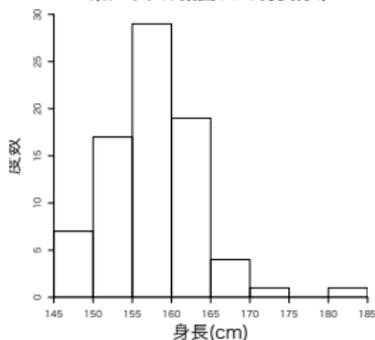
ここまで来たよ

- 1 はじめに
 - この授業どんなのり?
- 2 データの分布
 - データとは?
 - 度数分布表
 - ヒストグラム

ヒストグラム

高校 数学 I 西川確率統計 §5.1.1

某アイドル集団+1の身長分布



- ‘度数分布表を棒グラフにしたもの’
- 基本レベルでは階級幅は一定 \rightsquigarrow 本当は高さでなく面積
- 階級の個数や階級幅は指定がなければ、見やすいように決めてよい。
 - ▶ 階級の幅=超大きい \rightsquigarrow 長方形 1 個
 - ▶ 階級の幅=超小さい \rightsquigarrow ??

- ヒストグラムに限らず、グラフの縦軸横軸には量の名と単位を明示

L01-Q1

Quiz(度数分布表とヒストグラムを作ろう)

度数分布表とヒストグラムを手で作ろう. 全盛期のデータ.

名前	年齢						
		白間美瑠	17.6	入山杏奈	19.0	峯岸みなみ	22.0
		高橋朱里	17.1	生駒里奈	19.0	指原莉乃	22.0
中野郁海	14.1	向井地美音	17.0	木崎ゆりあ	19.2	横山由依	22.1
大和田南那	15.2	森保まどか	17.6	川栄李奈	20.7	松井玲奈	23.7
川本紗矢	16.4	松井珠理奈	18.1	武藤十夢	20.8	柏木由紀	23.2
大島涼花	16.2	渋谷凪咲	18.4	山本彩	21.4	須田亜香里	23.2
加藤玲奈	17.9	田野優花	18.4	島崎遥香	21.7	高橋みなみ	23.0
宮脇咲良	17.1	矢倉楓子	18.1	渡辺麻友	21.5	宮澤佐江	24.2
小嶋真子	17.9	兒玉遥	18.4	渡辺美優紀	21.2	小嶋陽菜	26.9

- 学籍番号奇数の人は 5 刻みで. 10-15,15-20,...,
- 学籍番号偶数の人は 4 刻みで. 12-16,16-20,...,
- 以上, 以下, 未満, より大きい, は自分で正しく決めて.

連絡

- 次回は 7-002 講義室
- 配布資料は 1-503 向かい掲示板前の引出,
<http://hig3.net> で再配布しています.
- 樋口オフィスアワー月 3.5(1-539) 金 4(1-502),
Math ラウンジ月-木昼 (1-614)
- 次回からは, 加減乗除と平方根 (ルート) の使
える電卓持ってきて. 関数電卓でなくても可.
スマホアプリでも可.
- Trial 予告



- 来週は教科書 西川確率統計 §5.1.2 代表値 §5.1.3 散らばりの尺度 読
んできて.
- 統計検定. 2017-11-26 一般受験, 団体割引希望
の人は樋口まで).