

受験者データの点数分布, 点数と順位

樋口さぶろお

龍谷大学工学部数理情報学科

生活の中の統計技術 L01(2018-09-24 Mon)

最終更新: Time-stamp: "2018-10-01 Mon 07:18 JST hig"

今日の目標

- ヒストグラムからもとのデータを想像できる.
- ヒストグラムから分位数を求められる
- ヒストグラムから箱ひげ図を作れる



ここまで来たよ

- はじめに
 - この授業どんなのり?

- 受験者データの点数分布, 点数と順位

科目の目標

もう少し正確にはシラバスを見てね.

- 2クラスの成績を比較できる
- クラスの成績の時間変化を比較できる
- 成績の違いの原因を探せる
- '正しい' アンケートが作れる
- '正しい' テスト問題が作れる

大学生のみなさんは生活の 1/3 くらいを学校で過ごすと思います. また, 学習塾でアルバイトする人や, 将来は学校で教育をしたいという人もいます. 学校ではテストやアンケートなどでデータが集められますが, これは, 学校による教育の状態, 学生/生徒の学習の状態を測定し, ある教え方がどの程度効果的か, 学生/生徒のある技能がどのくらい向上したか, などの問に答えるためのものです. 大学生のみなさんの慣れている, または想像が簡単な, 学校現場の具体的な状況を例として取り上げ, 問に答えるためにデータを取得して分析して統計的に結論を出す方法を説明します. 高等学校の数学 I 程度の予備知識を期待します. 詳細な数学的説明はせず, 意味と方法の説明を中心にします.

科目の位置づけ

- 統計を学校と教育の例で説明しますが, 他のことにも使えます.
 - ▶ 統計は〇〇科学の言葉.
 - ▶ 「データサイエンスは 21 世紀で最もセクシーな職業」 Harvard Business Review 2012
- 教育は教員になる人以外にも関係します.
 - ▶ 教育を提供する企業もあります.
 - ▶ ほぼすべての企業に企業内教育があります.

この内容を (教育をキープして) 進めると
教育統計学, 教育心理学, 教育経済学, 教育工学, ...

「学力」の経済学

統計学はときにセクシーな学問である

科目ののり

難しくありませんが、授業時間内は注文が多くめんどくさい科目です…
成績計算 科目の成績 100 ピーナッツは

- 40 ピーナッツ:平常点. 毎回授業での quiz, 授業時間外の予習復習.
- 30 ピーナッツ:プチテスト='中間試験'
- 30 ピーナッツ:紙のファイナルトリアル='期末試験'

現在の成績は e ラーニングサイトで見られるようになる予定.

欠席届 典型的には介護等実習

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). 何回欠席しても期末試験受験資格を失うことはありませんが、自分で追いついてね.

資料授業で配布. 授業後に欲しい人は <http://hig3.net> から各自ダウンロード. 1-503 前のレターボックスに残ってることも.

担当者ののり

- なまえ: 樋口さぶろお hig-stat@math.ryukoku.ac.jp
- へや: 1-502
- オフィスアワー: 火昼 (1-539), 金 14:40-15:40(1-502), 訪問歓迎な時間: 月金昼 (1-502).

Web ページ 資料配布 <http://hig3.net> → Moodle



テストのスコアレポートに載ってる情報から自分の位置を知ろう

スコアレポート

100 点満点

平均点 70 点

あなたの点数 68 点

自分の順位は? クラスの上半分? 下半分?

あるテストの結果

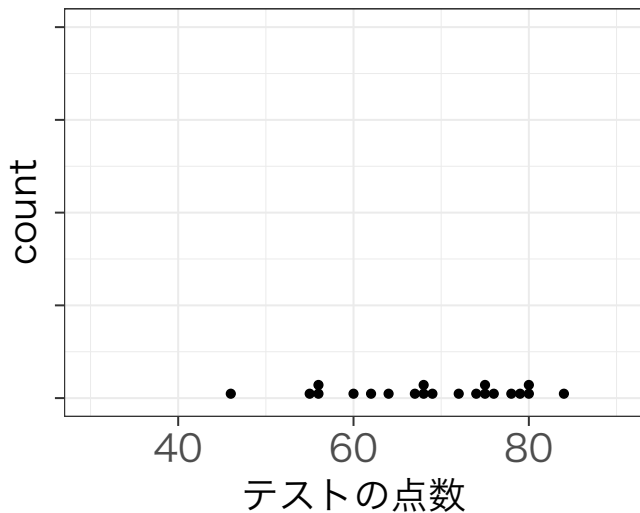
データの個数 = サンプルサイズ = サンプルの大きさ $N = 21$.

点数 (q 分位数)	順位	'順位'	q
46	1	0.0	0.000
55	2	1.0	0.050
56	3	2.5	0.125
56	3	2.5	0.125
60	5	4.0	0.200
62	6	5.0	0.250
...			
78	17	16.0	0.800
79	18	17.0	0.850
80	19	18.5	0.925
80	19	18.5	0.925
84	21	20.0	1.000

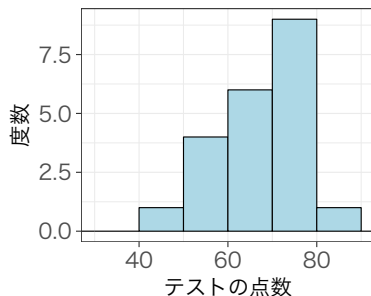
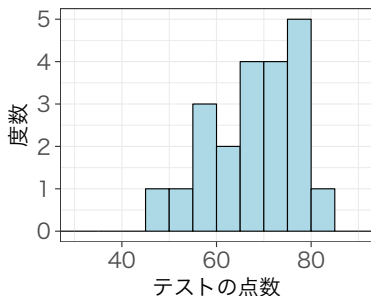
点数	'順位'	q
10	0	0.0
30	1	2.5
30	1	2.5
30	1	2.5
30	1	2.5
80	5	5.0

- このデータの 0.200 分位数は 60 点. 0.825 分位数は 78.5 点.
- $q = \text{'順位'}/(N - 1)$. q は
- とりあえず 四分位数とは別.
- q 分位数 = q -quantile, パーセントなら α -percentile

データをそのまま見る ストリップチャート



階級ごとに数える ヒストグラム



度数 = ($a \leq$ テストの点数 $< b$ の人数)

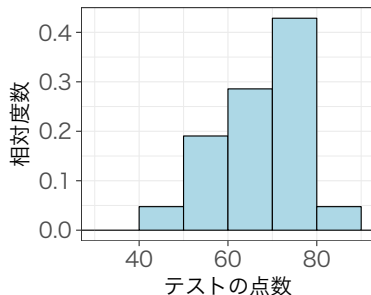
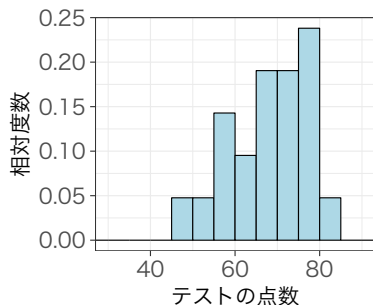
$[a, b)$ 階級

階級は, 人間が見やすいように決める. 合計は N .

順位が 4 の人の点数は?

62 点の人は何位?

相対度数で描いたヒストグラム



相対度数 = 度数 / 全体の人数. 合計は 1.00.

0.15 分位数は?

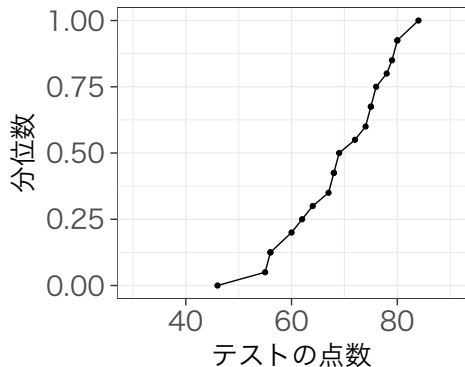
62 点は何分位数?

ふつうは度数分布表を経由する

階級	度数	相対度数
30 より大きく 40 以下	0	0.00
40 より大きく 50 以下	1	0.05
50 より大きく 60 以下	4	0.19
60 より大きく 70 以下	6	0.29
70 より大きく 80 以下	9	0.43
80 より大きく 90 以下	1	0.05
90 より大きく 100 以下	0	0.00
計	21	1.00

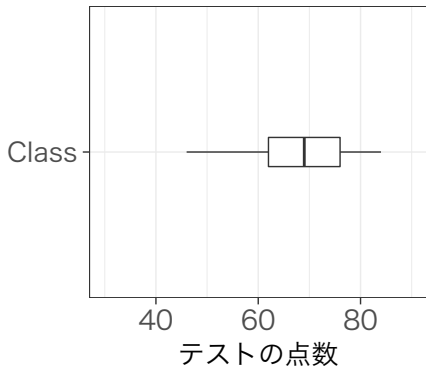
端はどちらの階級に入れるかはっきりしとけ?

データ-分位数の関係が ほぼ標本累積分布関数



ヒストグラム 対 データ-分位数関係
標本累積分布関数に似たものになってる.

四分位数だけ表示する箱ひげ図

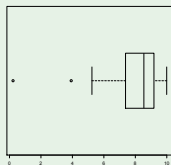
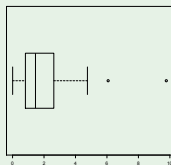
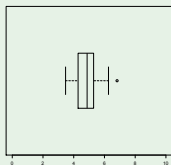
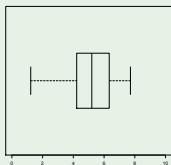
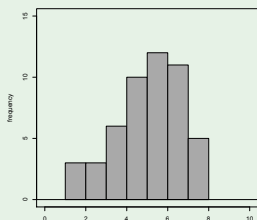


四分位数 = $0, 1/4, 2/4, 3/4, 4/4$ 分位数. ひげの先, 箱の両端と中央.

高校 数学 I の定義と少し違うが, N が大きいときは大差ない.

Quiz(ヒストグラムと箱ひげ図)

このヒストグラムに対応する箱ひげ図はどれ?



情報量

箱ひげ図 \geq ヒストグラム (階級幅大) $<$ ヒストグラム (階級幅小) $<$ ストリップチャート = 標本累積分布

情報 小 \leftrightarrow 大
見やすさ 大 \leftrightarrow 小

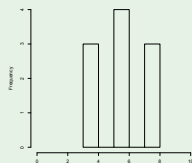
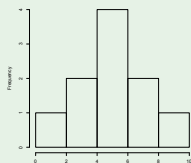
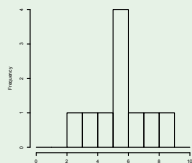
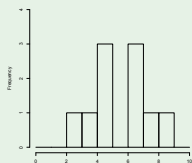
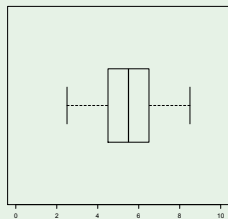
情報量の大小とは?

問

- ① 階級幅 5 のヒストグラムだけ見たとき, 0.25 分位数はいくつとわかる?
- ② ストリップチャートだけを見て, ヒストグラムで階級 $[50, 55)$ の高さはいくつかわかる?

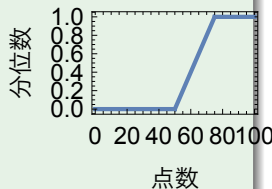
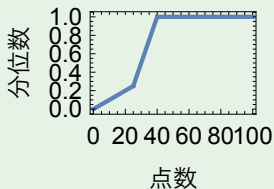
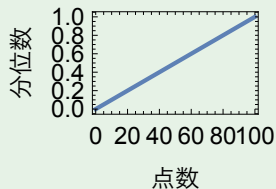
Quiz(ヒストグラムと箱ひげ図の対応)

この箱ひげ図に対応するヒストグラムはどれ?



Quiz(分位数とヒストグラム)

0 以上 100 以下の値を取る 3 組のデータがあり, データと分位数の関係がそれぞれ次のグラフのように表される. 各組のデータのヒストグラムを対比して描こう.



統計検定 3 級 (2017-06) から

2017 年 6 月統計検定 3 級問 11

統計検定 3 級 (2017-06) から

2017 年 6 月統計検定 3 級問 11

お知らせ

- 2018-10-01 月 2 はたぶん教室変更. 講義室か計算機実習室. ポータルで通知.
- 2018-11-25 日 統計検定の一般会場団体受験 (10%ディスカウント受験) (大阪会場) の 10 月初めまでに申込み. 申込

<https://goo.gl/forms/XUGUXr2TJw33wR3j1>

