

データの代表値と散布度

樋口さぶろお

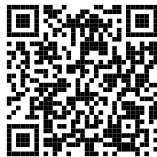
龍谷大学工学部数理情報学科

生活の中の統計技術 L02(2018-10-01 Mon)

最終更新: Time-stamp: "2018-10-01 Mon 17:02 JST hig"

今日の目標

- Excel でヒストグラムと箱ひげ図が描ける
- Excel で平均値と中央値, 分散と標準偏差を求められる
- Excel と手計算で1人の受験者の偏差値が求められる



ここまで来たよ

- 受験者データの点数分布, 点数と順位
 - Excel へのデータの読込とグラフ作成

- ① データの代表値と散布度
 - 代表値と散布度の意味

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

Excel でヒストグラム, 箱ひげ図を描こう 授業内課題 L02-1

- ① 動画 (の一部) Moodle からデータをコピー.

<https://learn.math.ryukoku.ac.jp>, Excel ワークシートに,
データの貼り付け, 区切り位置の調整. まず保存. ファイル名自由.

データ > 区切り位置の調整

- ② Excel で, データの領域を選択して, シート内にヒストグラムと 箱ひげ図を描く

メニュー > 挿入 > グラフ > ヒストグラム/箱ひげ図

- ③ Moodle に提出 動画

- ▶ 課題 → 課題を追加する
- ▶ ドラッグ・アンド・ドロップエリアにドラッグ
 - ★ または, 紙アイコンまたはドラッグ・アンド・ドロップエリアの下向きアイコンをクリック
 - ★ 長方形のアイコン (ファイル)
 - ★ ファイル選択
 - ★ このファイルをアップロードする
- ▶ 変更を保存する

グラフを描く他の方法

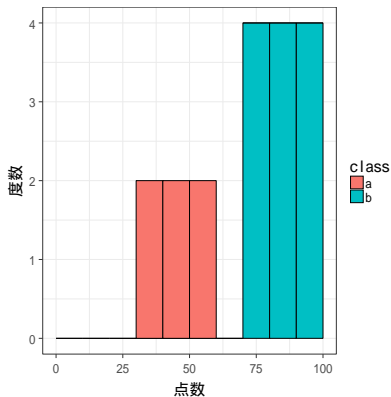
- Excel の枠内でも他の方法もあります.
- 社会 SPSS
- 社会/理工 R ggplot2
- 理工 Python matplotlib

ここまで来たよ

- 受験者データの点数分布, 点数と順位
 - Excel へのデータの読込とグラフ作成

- ① データの代表値と散布度
 - 代表値と散布度の意味

代表値



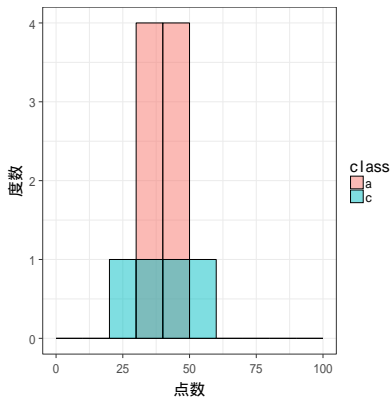
テストの点の代表値の違う2つのクラス.

代表値の違い

=

代表値にも何通りかある (後で)

散布度



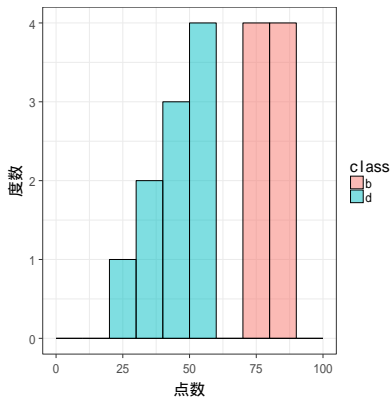
テストの点の散布度の違う2つのクラス.

散布度の違い

=

散布度にも何通りかある (後で)

いろんな代表値



平均値の意味

いいところ

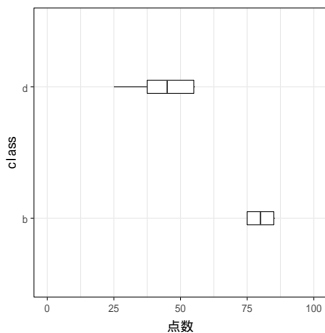
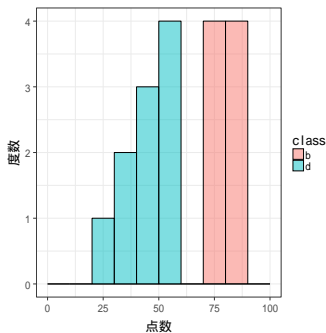
中央値の意味

いいところ

平均値 = mean = m 例 b $\frac{1}{8}(75 + \dots + 85) =$, 例 d

中央値 = median = 0.50 分位数 = , 例 d

いろいろな散布度



標準偏差 = $\sqrt{\text{分散}}$

分散 $\stackrel{\text{例 b}}{=} \frac{1}{8} [(75 - m)^2 + \dots + (85 - m)^2]$.

範囲 = (1.00 分位数) - (0.00 分位数) 意味

四分位範囲 = (0.75 分位数) - (0.25 分位数). 意味

いいところ

偏差値

偏差値

あるクラスの点数のデータの中の, 67 点 (の成績の人) の偏差値

$$50 + 10 \times \frac{67 - \text{平均値 } m}{\text{標準偏差 } s}$$

偏差値は代表値ではない

偏差値の平均値は , 偏差値の標準偏差は

Quiz(標準得点と偏差値を用いた比較)

数学と英語の学年共通のテストがあった. 数学の点数は 平均値は $\bar{x} = 90$, 分散 $S_x^2 = 16$ だった. 英語の点数は 平均値は $\bar{y} = 60$, 分散 $S_y^2 = 4$ だった.

- ① 数学で 86 点を取った人がいる. この 86 点に対応する標準得点と偏差値を求めよう.
- ② 数学での 86 点と英語での 57 点をくらべると, 学年内での相対評価として, どちらのほうがよい成績か. 式を使って理由を延べ, 完結した日本語の文で答えよう.

Quiz(偏差値)

学力テストの偏差値について、次のうち正しいものを1つ選ぼう。

- ① 偏差値の最低値は0である
- ② 自分の点が平均点より下のとき、分散が大きいほうが自分の偏差値はより高い
- ③ 自分の点が平均点より上のとき、標準偏差が大きいほうが自分の偏差値はより高い
- ④ 100点のテストで満点を取った場合の偏差値は75である
- ⑤ 偏差値50の人の順位は上から1/2である

Quiz(偏差値)

(学力) 偏差値について、次のうち正しいのはどれ(とどれ)?

- ① 偏差値の最低値は 0 である
- ② 偏差値の最高値は 75 である
- ③ 平均点 (をとった人) の偏差値は 50 である
- ④ 100 点のテストで満点を取った場合の偏差値は、他の人の成績しだいである
- ⑤ 偏差値 50 の人の順位は上から 1/2 程度である
- ⑥ 偏差値 60 の人の順位は上から 15% 程度である。

Excel で代表値, 散布度, 偏差値を求めよう 課題 L02-2

統計ツールを有効化 動画

ファイル > オプション > アドイン > Excel のアドイン > 設定 > 分析ツール に
チェックを入れて OK する.

統計ツール 動画

平均値, 標準偏差, 中央値

データ > データ分析 > 基本統計量 > 統計情報にチェック

分位数 データ > データ分析 > 順位と百分位数

注: 基本統計量で求めた分散を $\frac{n-1}{n}$ 倍, 標準偏差を $\sqrt{\frac{n-1}{n}}$ 倍すると, 今
考えている分散と標準偏差になる.

Excel 関数 これらも使える.

平均値 average

分散 var.p (var.s ではない. var.s は不偏標本分散)

標準偏差 stdev.p (stdev.s ではない. stdev.s は不偏標本標準偏差)

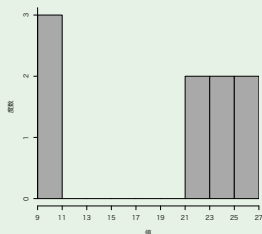
Moodle に書かれた課題をやって, アップロードして提出しよう.

L02-Q1

Quiz(平均値中央値最頻値)

次のヒストグラムから求めよう.

- ① 中央値
- ② (ヒストグラムの) 最頻値
- ③ 平均値



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

- 2018-11-25 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>

