

「データサイエンス」授業アンケート現状報告

2023/01/25

楊鯤昊

アンケート 内容

- 第1回・最終回主観アンケート
 - 9問：受講生が自ら情報リテラシー能力を評価
 - 最終回と第1回の評価を照らし合わせて、15回の授業を通して、（主観的に）能力が上がるかどうかを確認する。
 - $Score_{tn} - Score_{t0}$
- 毎回の授業評価アンケート
 - 出席の代わりに記入していただいている。
 - 毎回授業内容の難易度、授業への感想を確かめる。

主観アンケート

- 40人の学生数が第1回、第15回、両方の主観アンケートを記入した。
- 9割学生の評価が上がった。

この講義を後輩や他の学生に推奨したい。

個人情報、情報倫理、AI社会の原則等、数理・データサイエンス・AIの関連技術を活用する際の留意事項を理解している。

情報や知識を論理的に・多面的に分析することができる。

これまでに獲得した知識・技能・態度等を総合的に活用する力（実行力）を持っている。

新たな問題に直面する時、データ分析技術の視点からの問題解決策を見つける自信がある。

データ分析技術の視点から新たな問題を発見できる。

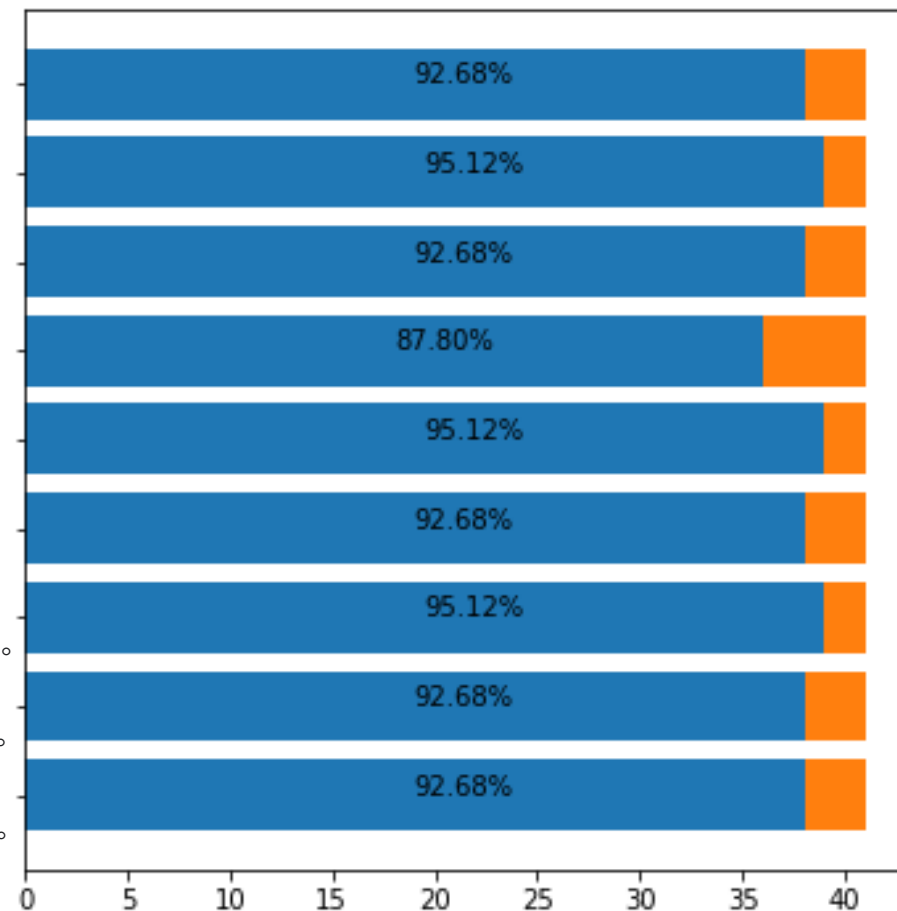
自分の専門において、AI技術がどのように応用されたかを理解している。

現実世界において、AI技術がどのように応用されたかを理解している。

コンピューターを使って必要な情報を収集し、それに基づいて資料を作成できる。

青色のバー：主観評価
が上がった学生数。

オレンジ色のバー：主観
評価が下がった学生数。



※パーセント：各ディメンションにおいて、主観
評価が上がった学生の比率

主観アンケート

平均的に、各ディメンションにおいて、2点ぐらい（0~10）評価が上がった。

⑨この講義を後輩や他の学生に推奨したい。

⑧個人情報、情報倫理、AI社会の原則等、数理・データサイエンス・AIの関連技術を活用する際の留意事項を理解している。

⑦情報や知識を論理的に・多面的に分析することができる。

⑥これまでに獲得した知識・技能・態度等を総合的に活用する力（実行力）を持っている。

⑤新たな問題に直面する時、データ分析技術の視点からの問題解決策を見つける自信がある。

④データ分析技術の視点から新たな問題を発見できる。

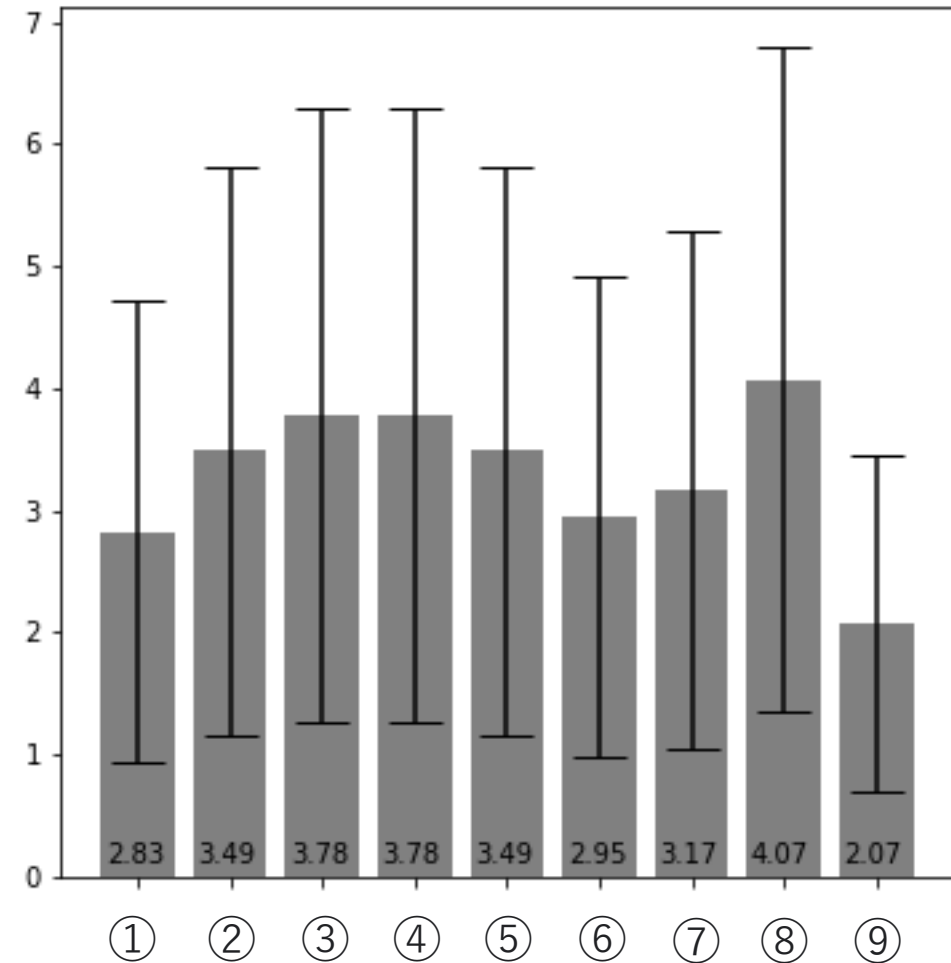
③自分の専門において、AI技術がどのように応用されたかを理解している。

②現実世界において、AI技術がどのように応用されたかを理解している。

①コンピューターを使って必要な情報を収集し、それに基づいて資料を作成できる。

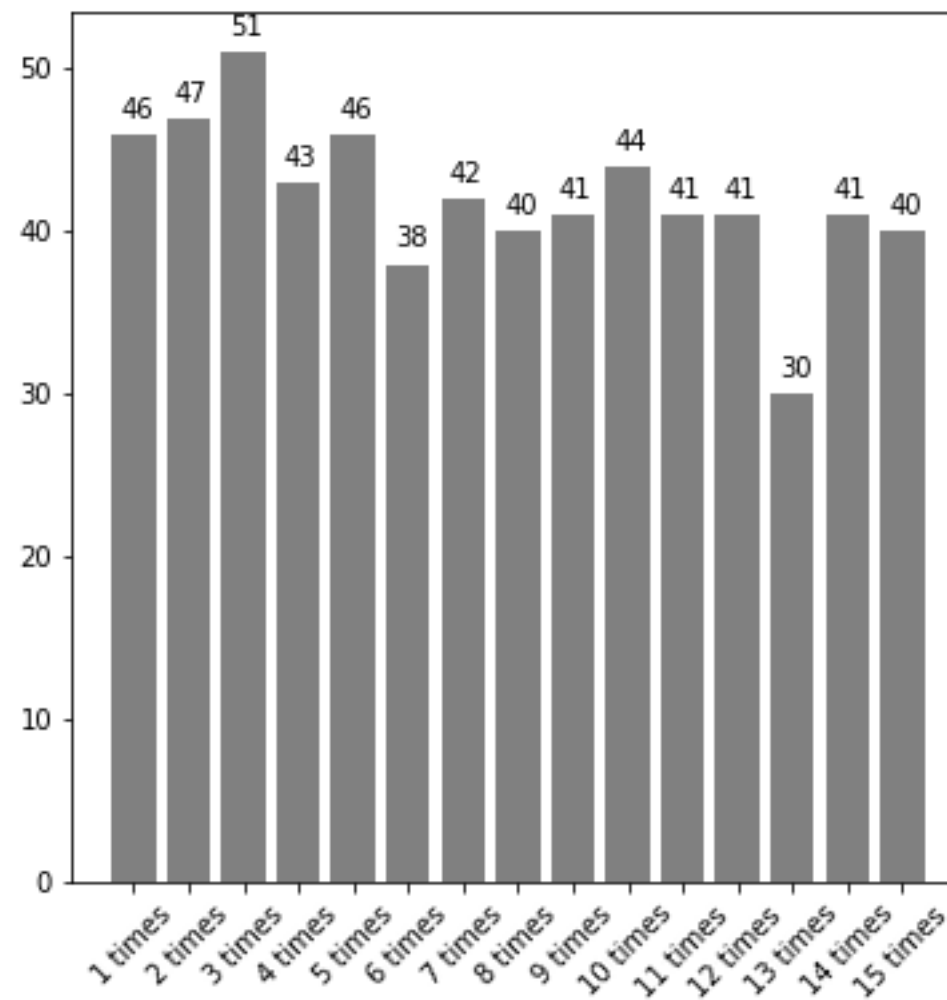
各ディメンション評価変動の平均値

※エラーバーは2回標準誤差を示す



出席状況

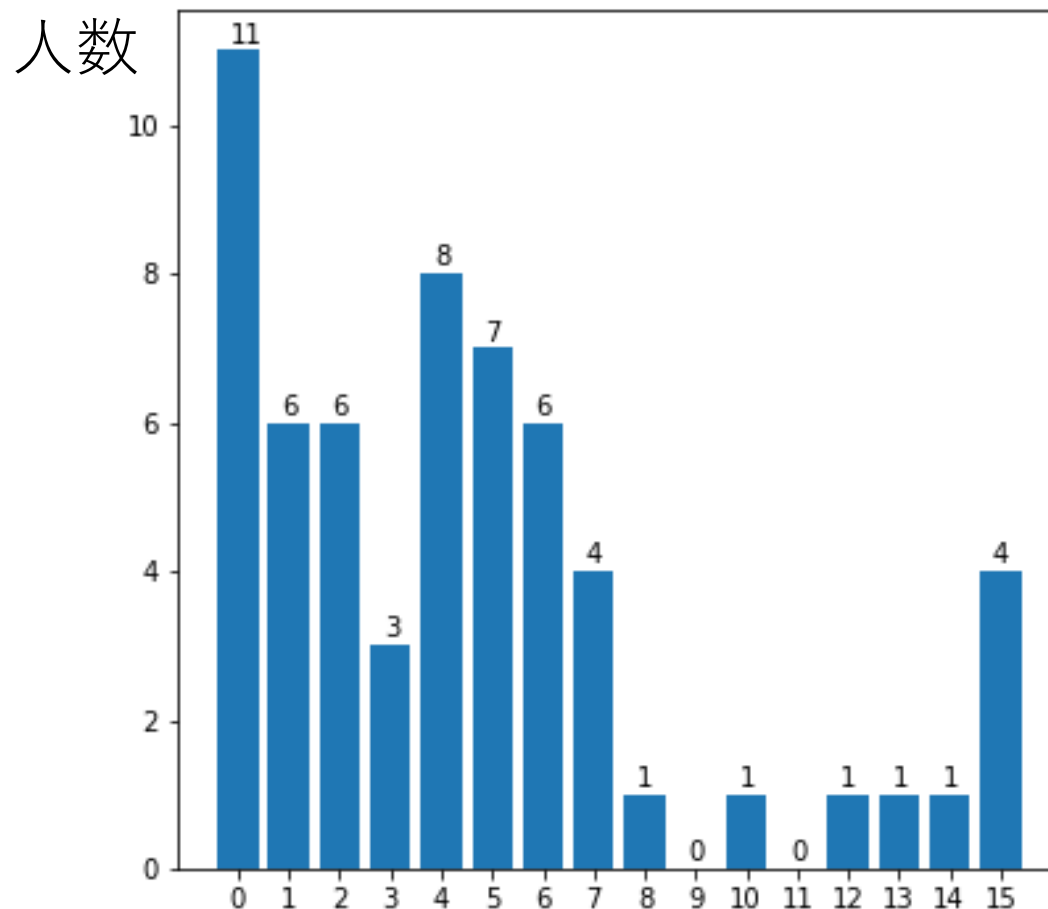
- 履修登録人数：60人
 - 15回にわたって、60%以上の受講生が出席してた。
 - ※出席を稼ぐための異常値（記入時間が異常、当学生の評価の時系列に従って評価点が異常）を除いている。
 - ※第13回（1月12日）の授業は楊がコロナ感染したため、オンライン式で行われたため、その一回の受講人数が少ない。



出席状況

- 個人別の出席状況を確認したところ、2割の学生が中途から脱退した（7回以上の授業を欠席）ことが判明した。
 - どうやって脱退率を下げるかは今後の課題だと考えられる（→山口先生にお伝えしました）。

学生欠席回数の分布



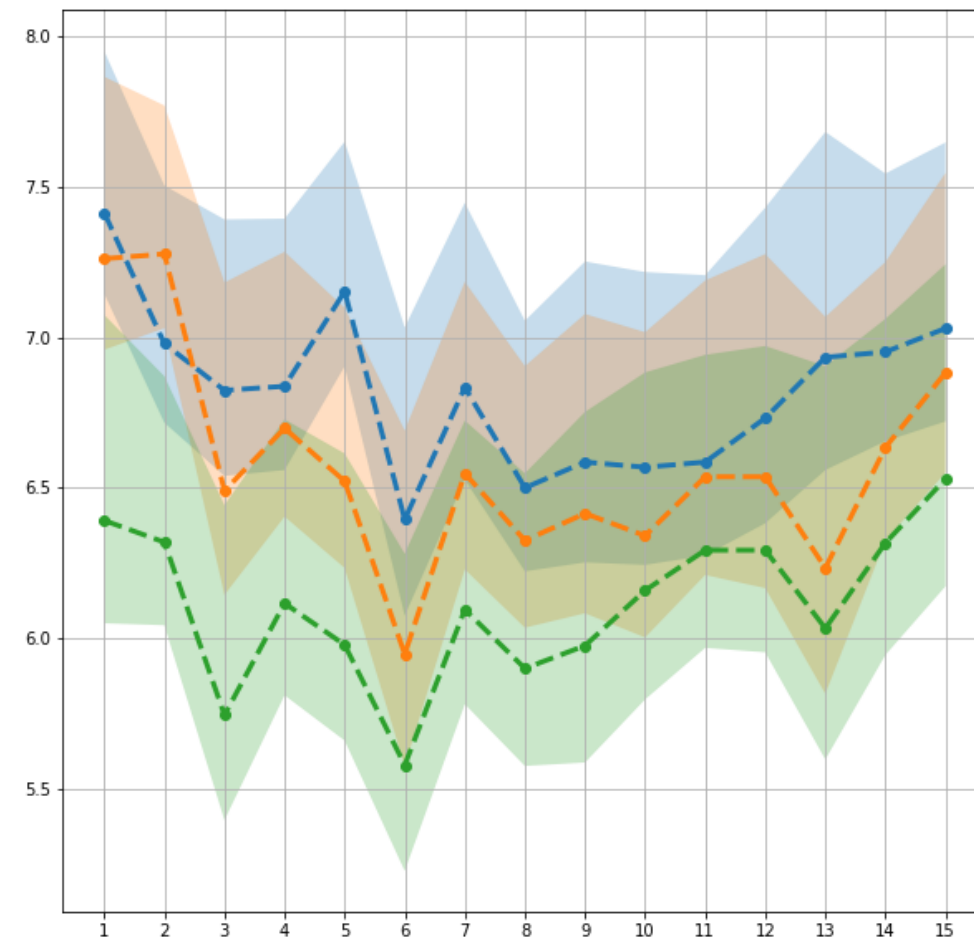
欠席回数の分布

平均評価

- 興味深さ、分かりやすさ、と理解度において、すべての授業の平均評価は**ポジティブ** (> 5) になっております。
 - ※色づきエリアは2倍標準誤差 ($\frac{\sigma(score_{ij})}{\sqrt{n_i}}$) を示す。
- その中、興味深さと分かりやすさにおいて、ほとんどの平均評価が6点を上回っている。
 - 「面白そうだけど、難しいところもあると感じた」というような感想を確かに多数いただいているので、今後の改善点として注力する必要がある。

平均評価
(0~10)

授業難易度の平均評価 (回数別)



授業回数

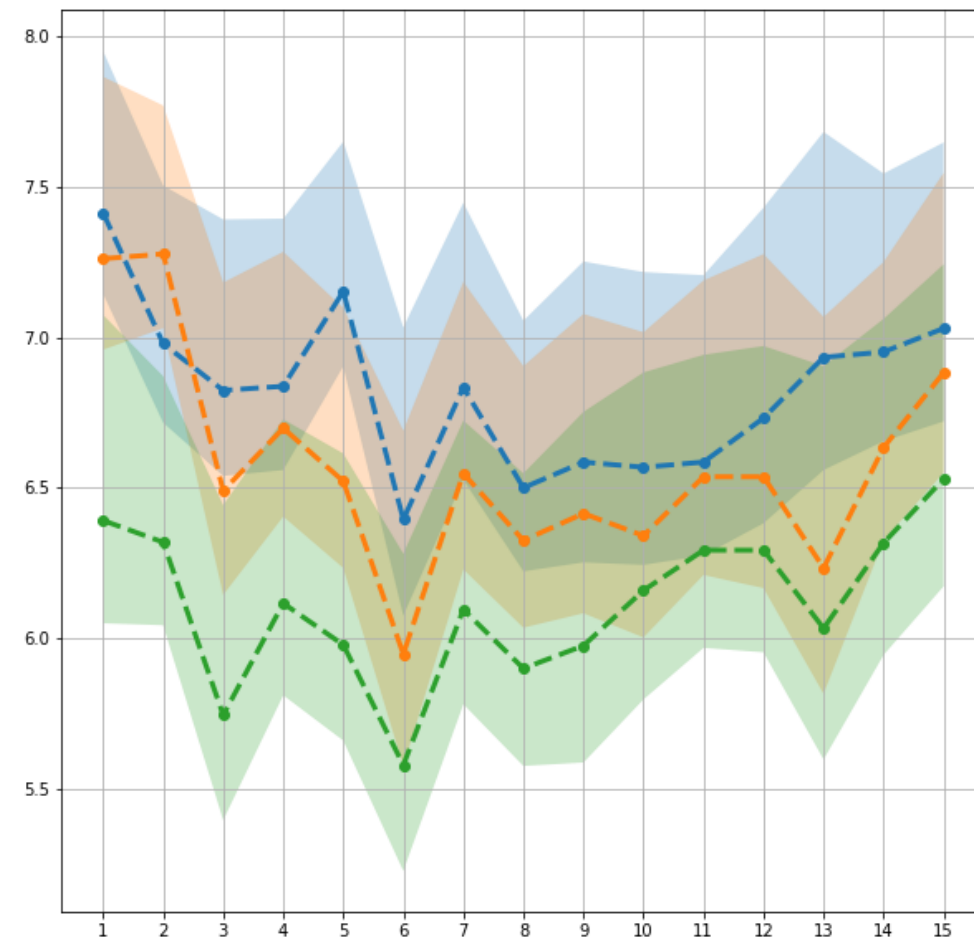
※青色は興味深さ、オレンジ色は分かりやすさ、緑色は自分の理解度における評価を示す

平均評価

- 興味深さ、分かりやすさ、と理解度において、すべての授業の平均評価は**ポジティブ** (> 5) になっております。
 - ※色づきエリアは2倍標準誤差 ($\frac{\sigma(score_{ij})}{\sqrt{n_i}}$) を示す。
- その中、**興味深さと分かりやすさ**において、ほとんどの平均評価が6点を上回っている。
 - 「面白そうだけど、難しいところもあると感じた」というような感想を確かに多数いただいているので、今後の改善点として注力する必要がある。

平均評価
(0~10)

授業難易度の平均評価 (回数別)



授業回数

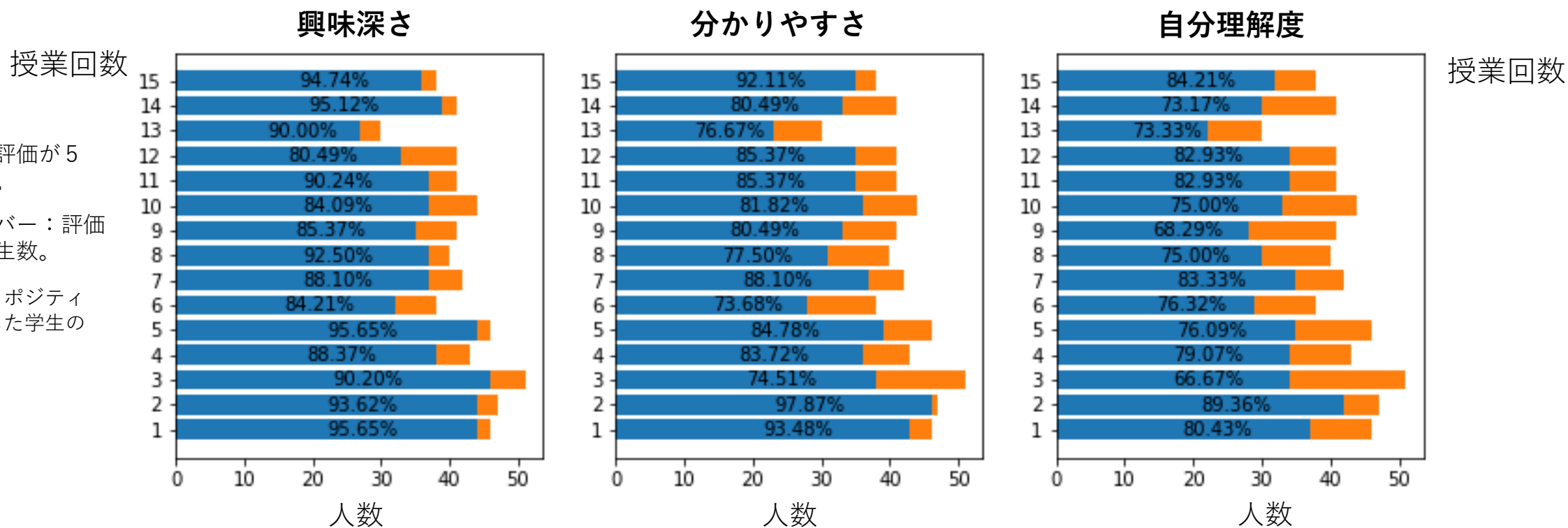
※青色は興味深さ、オレンジ色は分かりやすさ、緑色は自分の理解度における評価を示す

平均評価

- 難易度評価において、商学部所属の学生が高い評価を出す傾向性が見られた ($p\text{-value} < .001$) 。
- 法学部所属の学生と現代教養学部所属の学生の評価間に有意差が見られなかった。
- モデル：重回帰モデル
 - $Score \sim \beta_1 \cdot is_in_Business_Faculty + \beta_2 \cdot is_in_Education_Faculty + \beta_0$;
 - Where $is_in_Business_Faculty$ and $is_in_Education_Faculty$ were dummy variables

ポジティブvsネガティブな評価

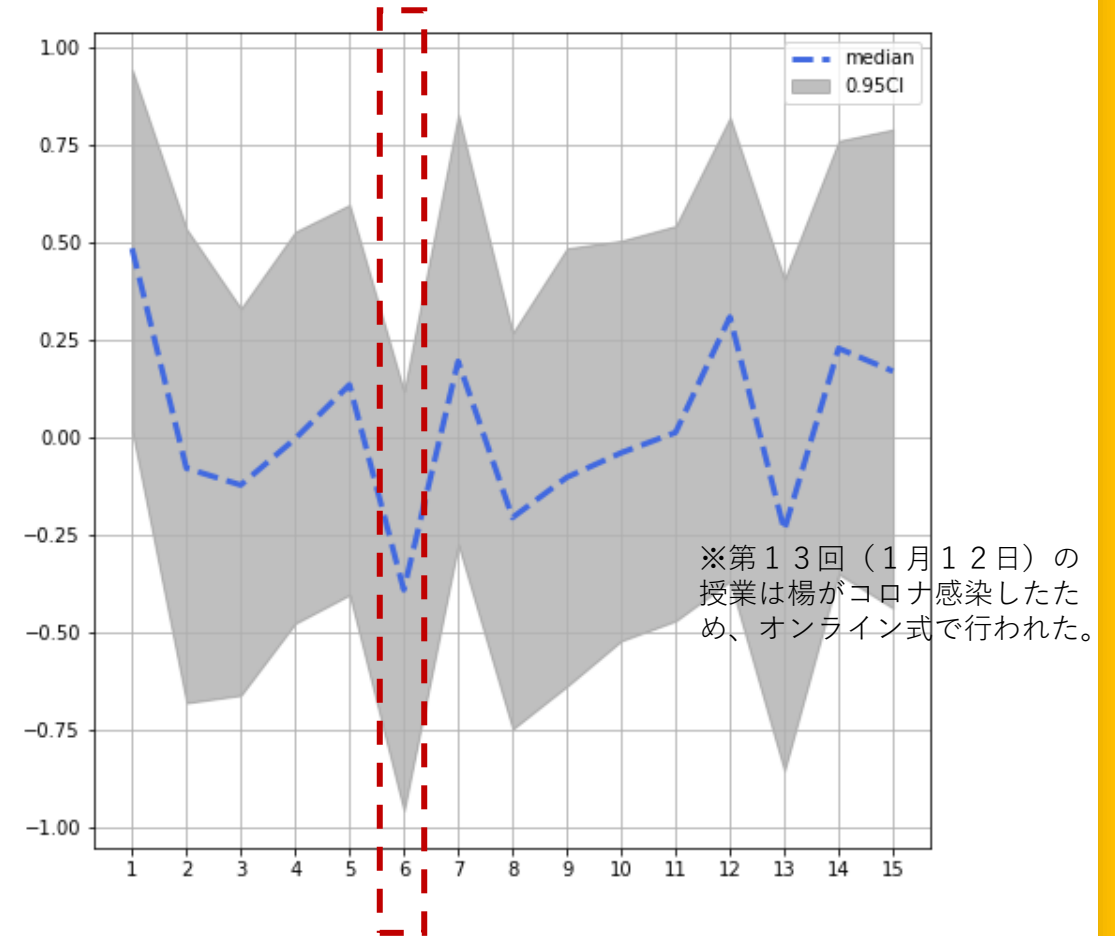
- 受講生の評価を「ポジティブ」(≥ 5)と「ネガティブ」(< 5)に分ける場合、興味深さ、分かりやすさ、と理解度において、**75%以上の受講生がポジティブな評価を出してくれた。**



授業改善のための総合モデル

- 授業の評価は時系列順で関連していると仮定
 - e.g., 第1回目のデータサイエンスの定義の説明が分からなかったら、第2回目データサイエンスとSociety 5.0の繋がりが分からなくなる。
- 興味深さ、分かりやすさ、理解度を総合的に考慮する。
 - e.g., 面白そうだけど、理解できていないことは意味がない。
- (連続の) 欠席は興味をなくした・挫折したを暗示している。
 - e.g., どうせい分からないから、聞いても意味ない。

- 第6回に「自然言語処理」と「画像処理」の原理を説明した。
- 反省点：導入が足りない上、(Word Embedding, Convolutionの原理に関する) 抽象的な話が多い。
- 改善策：原理の説明を軽くして、応用例を増やしたほうがいい。



まとめ

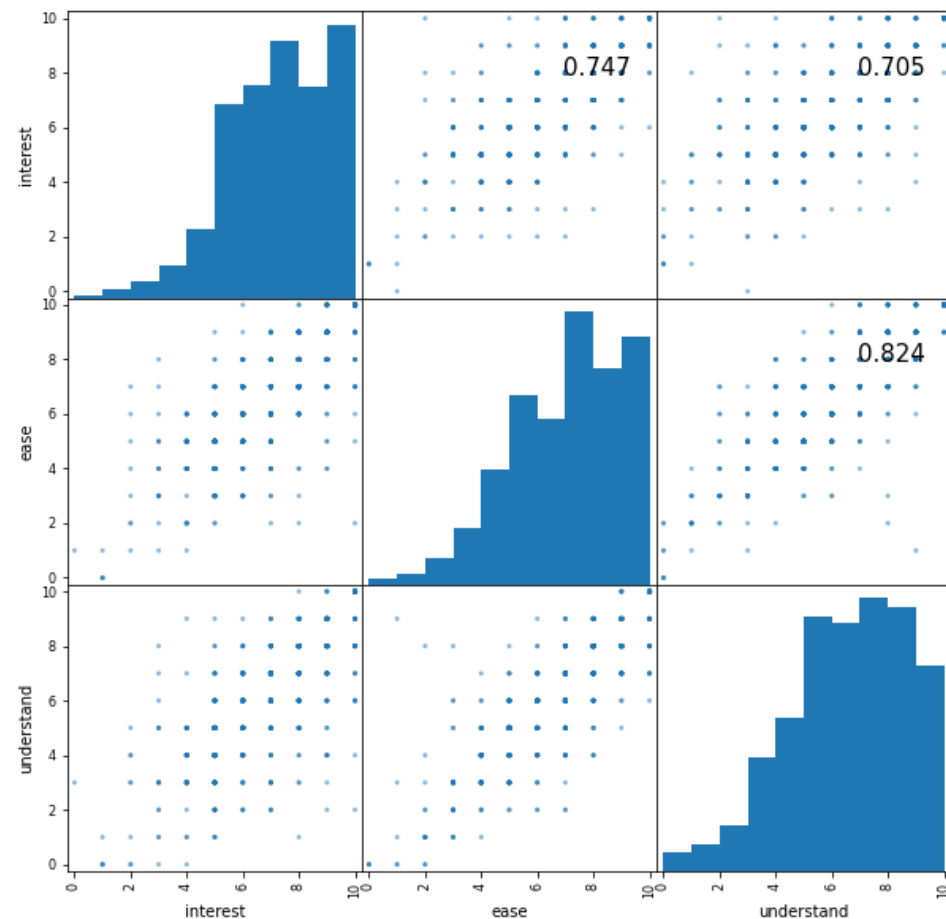
- 問題点：
 - データサイエンスに興味をもって登録した学生の評価が良かったが、興味がない学生の脱退率が高い。
 - 理解を深めるため、更に工夫が必要。

補足資料：授業改善のための総合モデル

- 隠れマルコフモデルを用いる。
 - NUTS法でパラメータを推定する。
 - すべてのパラメータのrhatが1.1より小さい（収束性）
 - 授業の出席状況を予測する時、 $R^2 = 66.42\%$ （精度）
 - 毎回授業の平均の総合評価を予測する時、 $R^2 = 85.17\%$ （精度）
- 授業の評価は**時系列順に関連**していると仮定
 - i 回目の授業の総合的な質を示す点数を α_i とする；
 - i 回目の授業への評価 s_i の分布は、 **i 回目まで**の授業の総合的な効果 β_i によって決まる。
 - $\beta_i = \alpha_i + \theta \cdot \beta_{i-1}$ ($\theta \in [0,1], \beta_0 = 0$)
 - $s_i \sim \text{StudentT}(\beta_i, \sigma, \nu)$

補足資料：授業改善のための総合モデル

- 興味深さ、分かりやすさ、理解度を総合的に考慮する。
 - 因子分析で、毎回の授業への総合的な評価 s_i を算出する。



補足資料：授業改善のための総合モデル

- (連続の) **欠席**は興味をなくした・挫折したを暗示している。
- **i回目まで**の授業の総合的な効果 β_i は**i+1回目**の出席人数 ad_i に影響している。
 - $prob_i = \phi\left(\frac{\gamma_0}{i} + \gamma_1\beta_i + \gamma_2\right)$; $ad_i \sim \text{Binominal}(prob_i, 60)$
 - where $\phi(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function
 - when $prob_i = \phi\left(\frac{\gamma_0}{i} + \gamma_2\right)$

