



EDAを効率化するための 汎用データ例

山形大学 学術研究院 教授

藤原 宏司

IR初級者セッション（大学評価・IR担当者集会2023）

2023.8.31 @ JR博多シティ

「Data250_20230831.xlsx」のイメージ

ID	性別	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15	成績素点
00271	女性	出席	出席	出席	欠席	出席	出席	出席	出席	欠席	出席	出席	出席	欠席	出席	出席	73
01144	女性	欠席	出席	欠席	出席	欠席	出席	出席	欠席	出席	69						
00517	男性	出席	欠席	出席	欠席	出席	出席	欠席	出席	出席	欠席	欠席	出席	出席	欠席	出席	33
01364	男性	出席	出席	出席	出席	出席	欠席	欠席	出席	欠席	欠席	欠席	出席	欠席	出席	欠席	38
00389	男性	欠席	出席	出席	欠席	出席	出席	出席	欠席	欠席	出席	出席	欠席	出席	欠席	欠席	46
00968	男性	出席	出席	出席	出席	欠席	出席	欠席	出席	欠席	出席	欠席	出席	出席	欠席	出席	65
00856	女性	出席	出席	欠席	出席	欠席	出席	欠席	出席	欠席	84						
00724	女性	出席	欠席	出席	出席	出席	出席	欠席	欠席	出席	欠席	出席	欠席	欠席	欠席	出席	34
00811	男性	欠席	欠席	欠席	欠席	欠席	出席	出席	出席	出席	出席	欠席	出席	欠席	出席	欠席	42
00738	男性	出席	欠席	欠席	出席	出席	出席	欠席	出席	出席	出席	出席	欠席	欠席	出席	欠席	40
00723	男性	欠席	出席	出席	出席	出席	欠席	出席	出席	欠席	出席	欠席	出席	出席	欠席	出席	70
00346	女性	欠席	出席	出席	出席	出席	出席	欠席	出席	欠席	出席	出席	出席	出席	欠席	欠席	62
01072	女性	出席	欠席	出席	出席	出席	欠席	出席	出席	出席	出席	出席	出席	欠席	出席	出席	72
00790	その他	欠席	出席	出席	出席	欠席	出席	欠席	欠席	出席	欠席	欠席	欠席	出席	出席	欠席	43

■ ID = 学生番号

C01 = とある科目における初回の出欠状況

C15 = 最終回の出欠状況

成績素点は100点満点

■ N = 250

【再掲】 問いを考えてみよう

■ 定番と呼べる「問い」の例 ■

1. 「出席率」 vs. 「成績」
2. 「出席率」 vs. 「性別」
3. 「成績」 vs. 「性別」
4. 「欠席回数」 vs. 「単位取得状況」

午後の演習問題

■ 少し複雑な「問い」の例 ■

1. 「**3回以上連続欠席**の有無」 vs. 「成績」 vs. 「性別」
2. 「**連続欠席複数回**の有無」 vs. 「成績」 vs. 「性別」
3. 「連続出席の回数」 vs. 「成績」 vs. 「性別」
4. 「連続欠席が無い学生」における「欠席回数の分布」

汎用的な分析用データの作成

■ なぜ、そのようなデータが必要？

→ 「問い」ごとにデータを作り直したり、
カウントをやり直したりするのは**面倒、苦痛、時間の○○**

例：2回連続欠席、3回連続欠席、10回以上連続出席、etc.

似たような問い

→ 上記のような問いに「まとめて」対応できれば楽

→ 汎用性のある分析用データの作成へ

Run Length Encoding (RLE)

ID	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
00724	出席	欠席	出席	出席	出席	出席	欠席	欠席	出席	欠席	出席	欠席	欠席	欠席	出席

- ID = 00724の学生さんは、

出席1回、**欠席**1回、出席4回、**欠席**2回、出席1回、**欠席**1回
出席1回、**欠席**3回、出席1回

という出欠パターンでした。

ID	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
XXXXX	出席	欠席	出席												

- ID = XXXXXの学生さんは、

出席1回、**欠席**13回、出席1回

という出欠パターンになります。

この方法を使えば、
出欠状況を授業回ごと
ではなく、**圧縮した状態で
表現可能**

RLEを応用した作成したデータ例

■ ID = 00724のオリジナルデータ

ID	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
00724	出席	欠席	出席	出席	出席	出席	欠席	欠席	出席	欠席	出席	欠席	欠席	欠席	出席

■ 出欠パターン

出1回、**欠1回**、**出4回**、**欠2回**、出1回、欠1回、出1回、**欠3回**、出1回

■ 汎用性を持たせたデータ例 → 「フィルター処理」で対応が容易になる

ID	性別	出席状況	授業開始回	授業終了回	連続回数
00724	女性	出席	1	1	1
00724	女性	欠席	2	2	1
00724	女性	出席	3	6	4
00724	女性	欠席	7	8	2
00724	女性	出席	9	9	1
00724	女性	欠席	10	10	1
00724	女性	出席	11	11	1
00724	女性	欠席	12	14	3
00724	女性	出席	15	15	1



■ 「Data250_20230831v2.xlsx」を参考に試してみましよう！

THANK YOU!

ANY QUESTIONS, COMMENTS OR SUGGESTIONS?

藤原 宏司 | Koji Fujiwara, Ph.D.

kfujiwara@cc.yamagata-u.ac.jp

