

大学評価・IR担当者集会 2023 IR初級者セッション

Excel による分析例

— 演習まとめ：IR担当者としてのNext Step —

主催：大学評価コンソーシアム、九州大学 基幹教育院 次世代型大学教育開発センター
2023年8月31日（金）JR博多シティ

山本 幸一（明治大学）

https://researchmap.jp/ko_yamamoto/presentations/28302618

出欠データのカウントはできましたか？ 難しかった？

- この演習で大切なこと
正解できたこと、Excelやツールが使えたこと、
では、ありません!!



理解いただきたかったこと、とは？

- IR担当者として、**ローデータに触れること**の大切さ
- 分析では、定番・常識に囚われず、**さまざまな切り口**を試みて、
データの特徴を捉えること
- そのためには、データ分析に適した**データの持ち方**や、**データマネジメント（前処理、加工等）**の知見が必要になること

Excelによる分析例から、演習のポイントを確認しましょう





生成系AIの分析例をきっかけにIR担当者のスキルを考える

- 生成系AI…IR担当者として**結果を検証できるデータ分析の力量**は必要では？
- ある問題に正解しても、次の質問も正解できるかは、聞き方次第。聞き方、すなわち、**どのように分析用のデータを作成（前処理、加工、蓄積、再利用等）**するのかという知識はIR担当者として必要では？



データ分析経験のある人は、**汎用性の高い分析用データ**を作り、**そのデータに様々な問いを投げかける分析**を行っている。
(探索的データ分析：EDA)



① 何となく分析（ex.データの相関だけ）していませんか？

- Pivot（Excel）、BIツール（power BI、Tableau…）等では、簡単に分析らしいことができる。クロス集計、相関分析…。
- しかし「Data-based decision making」というように、IRの問い（分析テーマ、RQ）は、“**データに基づくもの**”。つまり、様々な切り口でデータ分析してこそ、本当に調査すべき問いを見出すことができる。
- データ内容を深く理解せず、BIツールにデータを投入し「何となく」相関を確認するだけでは、データに基づく分析になっていないかも知れない。

データ演習のポイント①

- 「決めつけ」で分析せず、様々な切り口を探して、データ分析してみよう。
- このデータを理解する作業（探索的データ分析）があつてこそ、データに基づく分析といえ、その結果、良いRQ（分析テーマ）に繋がる。



② 例えば、出欠データを“探索的にデータ分析”する

- 分析の定番は、グルーピング（○○ごとの△△）して比較。
 - 学科ごと×欠席率、学科ごと×性別ごと×欠席率
 - 欠席率と成績との相関 … BIツール等で簡易にできる
- この他に、グルーピングできないか（データの特質を探る）。
 - 3回以上連続欠席の有無×成績×性別 ← 今回課題
 - 5回目までに2回以上欠席の有無×成績
 - 7回目までの欠席回数より8回目以降の欠席回数が多い学生×成績
 - 成績80点以上と60点未満の学生×欠席回数の分布 …

データ演習のポイント②

どこに、どのような、特徴があるのか分からないデータも多いので、さまざまな角度からデータを確認する（EDA:探索的データ分析）。



- 問1 「3回以上連続欠席」の人数/平均点/性別ごと
 - ① **TEXTJOIN (値の連結)** 1・1・1→111
 - ② **COUNTIF** 欠席=0と置換し、*000* をカウント
 - ③ 人数…COUNT系
 - ④ 平均…AVERAGE系

- 問2 「連続欠席が複数回」の人数/平均点/性別ごと
 - ① ワイド型を、ロング型に、**データ型の変換** (Pivot解除)
 - ② **TEXTJOIN** 学生ID+出欠記号でユニーク値を作成
 - ③ **RLE (Run Length Encoding)** : 連続する同一の値を列の長さの値に変換 (置換)
 - ④ **IF系** : 連続する同一値をカウントし最大値を表示
 - ⑤ **Pivot** : 欠席データを対象、連続回数1及び空白は除外、総計が2以上をカウント



ワイド型データ

- 1行に複数の観測値をもつデータ
- “目視”では、直感的に理解しやすい
- 1件1行の横に広い形式

受験番号	性別	国語	外国語	数学	世界史	日本史	合計
1001	女性	75	82			58	215
1002	男性	68	62	86			216
1003	男性	82	78		71		231
1004	女性	85	87	78			250

ロング型データ

- 1行に1つのみ観測値をもつデータ
- データ分析で扱いやすい
- 1件複数行の縦長の形式



受験番号	性別	科目	点数
1001	女性	国語	75
1001	女性	外国語	82
1001	女性	日本史	58
1002	男性	国語	68
1002	男性	外国語	62
1002	男性	数学	86
1003	男性	国語	82
1003	男性	外国語	78
1003	男性	世界史	71
...

データ演習のポイント③

適切に分析するには、データマネジメント、前処理（空白処理、データ変形・結合等の加工等）のスキルが必要になる。

TEXTJOIN

T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
ID	性別	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15	成績素点		TEXTJOIN
00271	女性	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	73		111011110111011
01144	女性	0	1	0	1	0	1	1	0	1	1	1	1	1	1	1	69		0101011011111111
00517	男性	1	0	1	0	1	1	0	1	1	0	0	1	1	0	1	33		101011011001101
01364	男性	1	1	1	1	1	0	0	1	0	0	0	1	0	1	0	38		111110010001010
00389	男性	0	1	1	0	1	1	1	0	0	1	1	0	1	0	0	46		011011100110100
00968	男性	1	1	1	1	0	1	0	1	0	1	0	1	1	0	1	65		111101010101101
00856	女性	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0	84		110101111111010
00724	女性	1	0	1	1	1	1	0	0	1	0	1	0	0	0	1	34		101111001010001
00811	男性	0	0	0	0	0	1	1	1	1	1	0	1	0	1	0	42		000001111101010
00738	男性	1	0	0	1	1	1	0	1	1	1	1	0	0	1	0	40		100111011110010
00723	男性	0	1	1	1	1	0	1	1	0	1	0	1	1	0	1	70		011110110101101
00346	女性	0	1	1	1	1	1	0	1	0	1	1	1	1	0	0	62		011111010111100
01072	女性	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	72		101110111111011
00790	その他	0	1	1	1	0	1	0	0	1	0	0	0	1	1	0	43		011101001000110
00907	男性	0	1	1	1	0	0	1	0	1	1	1	1	1	0	0	43		011100101111100
00108	女性	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	72		111110111011101
00746	女性	1	0	1	0	1	0	1	0	0	1	0	1	1	1	1	32		101010100101111
00459	女性	0	1	0	1	1	1	1	0	1	0	1	1	1	1	0	68		010111101011110
01491	男性	1	0	0	1	1	0	1	0	1	1	1	1	1	1	1	74		100110101111111
00961	男性	0	1	1	0	1	1	0	1	0	0	0	1	1	1	1	51		011011010001111
00258	男性	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	97		111111101111111
01451	女性	1	0	1	0	1	1	1	0	1	1	1	0	1	0	0	28		101011101110100
00749	男性	0	0	0	1	1	1	1	0	1	0	0	1	0	0	0	23		000111101001000
00026	男性	1	1	1	0	0	1	1	0	0	1	1	0	0	0	1	42		111001100110001
01179	女性	0	1	0	1	1	0	0	0	1	1	1	0	1	1	1	38		010110001110111
01077	男性	1	1	0	0	1	1	0	1	1	1	1	1	0	1	1	68		110011011111011
00212	女性	1	0	0	1	1	1	1	1	1	0	1	0	1	1	1	85		100111111010111
00359	男性	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	75		111011101111111
00629	女性	1	1	1	1	0	0	1	0	0	1	0	1	0	1	1	33		111100100101011
00445	男性	1	1	1	0	0	1	1	0	1	0	0	1	1	1	0	48		111001101001110
01046	男性	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	90		111110011111111
00031	男性	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	83		111110111101110

Run Length Encoding

	A	B	C	D	E	F	G	H
1	ID	性別	成績素点	授業回	出1/欠0	記号(ID+出欠)	記号2	連続回数計
2	00271	女性	73	C01		1 002711	1	
3	00271	女性	73	C02		1 002711	2	
4	00271	女性	73	C03		1 002711	3	3
5	00271	女性	73	C04	0	002710	1	1
6	00271	女性	73	C05		1 002711	1	
7	00271	女性	73	C06		1 002711	2	
8	00271	女性	73	C07		1 002711	3	
9	00271	女性	73	C08		1 002711	4	4
10	00271	女性	73	C09	0	002710	1	1
11	00271	女性	73	C10		1 002711	1	
12	00271	女性	73	C11		1 002711	2	
13	00271	女性	73	C12		1 002711	3	3
14	00271	女性	73	C13	0	002710	1	1
15	00271	女性	73	C14		1 002711	1	
16	00271	女性	73	C15		1 002711	2	2
17	01144	女性	69	C01	0	011440	1	1
18	01144	女性	69	C02		1 011441	1	1
19	01144	女性	69	C03	0	011440	1	1
20	01144	女性	69	C04		1 011441	1	1
21	01144	女性	69	C05	0	011440	1	1
22	01144	女性	69	C06		1 011441	1	
23	01144	女性	69	C07		1 011441	2	2
24	01144	女性	69	C08	0	011440	1	1
25	01144	女性	69	C09		1 011441	1	
26	01144	女性	69	C10		1 011441	2	
27	01144	女性	69	C11		1 011441	3	
28	01144	女性	69	C12		1 011441	4	
29	01144	女性	69	C13		1 011441	5	
30	01144	女性	69	C14		1 011441	6	
31	01144	女性	69	C15		1 011441	7	7
32	00517	男性	33	C01		1 005171	1	1



① 繰り返し行うことが難しい

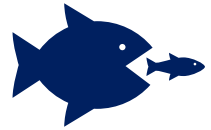
- 途中でデータ内容に変更があった場合（出欠データに修正があった、1名減った等）、最初からやり直し。
- 同じような（科目だけ異なる等）依頼があっても、再現する場合には、最初から組み直し。

② 分析手順を残しておくことが難しい

- 作成者が休暇中の代替や、人事異動による引継ぎで、間違いなく実行するには、かなり困難を伴う。

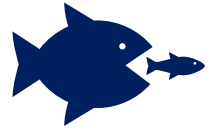
今回のデータ演習のポイント④

- 今回の演習は1科目。実践では、10～100科目と膨大になる。データを「再現」できることが、効率よく（楽しんで）IRを行う近道。
- 将来的には、データマネジメント（R等）のスキルも。



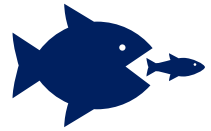
① ローデータに触れる！

- データの内容、中身、性質、性格を理解する。IR担当者として重要なこと。



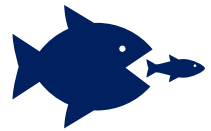
② 定番・常識に囚われない 様々な分析（探索的データ分析：EDA）

- どこに、どのような、特徴があるのか分からないデータも多い。
- データの特質を理解して分析することが“データに基づく分析”。
- 何となくデータを使っただけの分析は、データに基づく分析ではない場合も。



③ データマネジメント

- データ前処理（加工、データ変形・結合等）のスキルをもつ。
- ワイド型をロング型に変形する（分析できるように変形する）。



④ データの再現性を確保する

- 学科だけ異なる同じレポートの作成、楽しんで大量のデータ処理をするスキルは？
- Excelだけではなく、再現性あるツール（R等）の活用



Making Campus Decisions Better



Thank you for your attention.

明治大学 山本 幸一 ma94004@mics.meiji.ac.jp