



So, we have the data. Now what?

～探索的データ分析とリサーチクエスチョン～

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

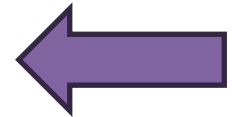
今日の内容

■ 自己紹介（次のページ）

■ 今回のテーマ

IR担当者が抱える代表的な悩み事として、

1. データが入手できない
2. データはあるけど、何をすれば良いか分からない



といったものがあります。今回は、上記「2番」に着目します。午後の演習問題を想定しながら、

「探索的データ分析」と「リサーチクエスション」の関係等について、一緒に考えていきましょう。



（持ち時間：50分）

今日の内容 (contd.)

■ 自己紹介：藤原 宏司 | Koji Fujiwara

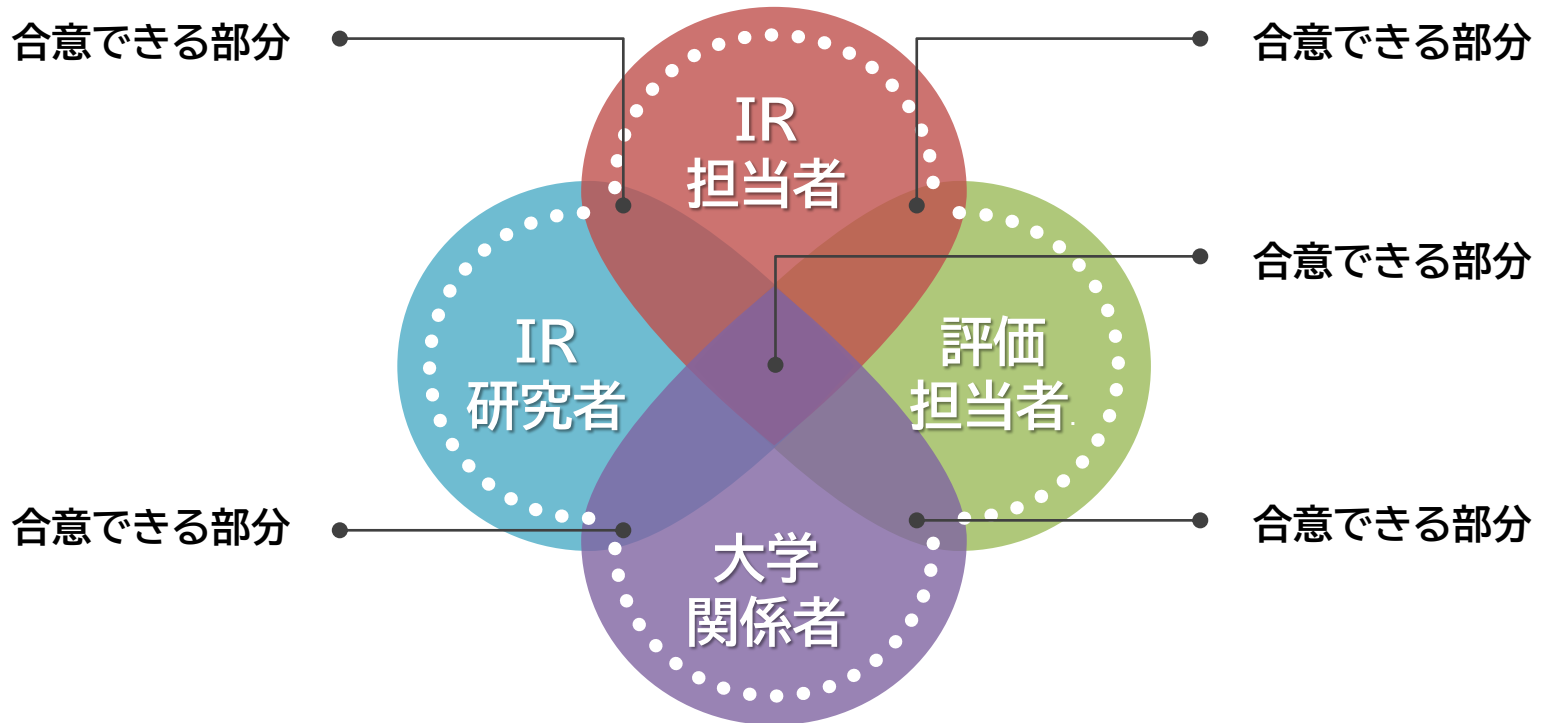
- 専門は、**米国における高等教育とIR (Institutional Research)** および応用統計学 (Ph.D.)
- 米国の大学・短期大学等で、データ分析、IRおよび大学評価対応業務に従事
- 2016年8月から現職
- 「山形大学IR担当者向け実践プログラム」プログラムディレクター

【注意点】発表者の考えるIRとは？

- **IR** (Institutional Research) とは、**アメリカの大学**で**発展**してきた、**計画立案、政策形成、意思決定等のサポート**に資する**情報提供**を目的に、**学内外データを収集・可視化する機能** (活動)
- **IRは「サポート」機能**
 - クライアントに分かりやすい正確な**レポート**を提供する **迅速に**
 - **【間違い】** 改善に役立つ「何か」を探して、提供しなければサポートではない 
 - **【目標】** 「IRのデータが意思決定の参考になった」(こう言われるように仕事ができれば…) 

日本の場合 → 立場によって考え方が異なります

■ IRとは？ IR担当者に期待される役割とは？



- 「IR担当者」「評価担当者」「IR研究者」「大学関係者」の間でも、考え方は当然異なります。→ **自分で注意深く考えよう！**

- **参考：米国の大学において、IRの役割を定義するのは、その大学の執行部**
 - IR担当者等は、執行部のオーダーに従うのが業務上の義務
 - IR研究者が持つ「大学関係者」に与える影響力は、ほぼ皆無

So, we have the data.

Now what?



シナリオ

■ 教学担当副学長から、

とある科目の出席データと科目成績（次ページ参照）
を渡され、

「このデータを分析して欲しい」

と依頼されました。

■ ポイント

- 分析上の具体的な指示はありませんでした。（あるある？）

データのイメージ（とある科目の出欠と成績）

ID	性別	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15	成績素点
00009	女性	出席	欠席	出席	欠席	出席	出席	出席	出席	出席	出席	出席	欠席	出席	欠席	欠席	73
00067	男性	欠席	出席	欠席	欠席	出席	欠席	欠席	出席	出席	欠席	出席	出席	欠席	出席	欠席	41
00136	女性	欠席	欠席	欠席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	欠席	39
00161	女性	欠席	欠席	出席	欠席	欠席	出席	出席	出席	出席	出席	出席	出席	出席	欠席	欠席	45
00171	女性	出席	欠席	出席	出席	欠席	欠席	出席	欠席	出席	欠席	出席	出席	欠席	欠席	出席	42
00186	女性	出席	出席	出席	欠席	出席	出席	欠席	出席	欠席	出席	出席	出席	出席	出席	欠席	69
00201	男性	欠席	出席	出席	欠席	出席	出席	出席	欠席	出席	出席	出席	欠席	出席	欠席	出席	65
00236	女性	出席	出席	欠席	欠席	出席	出席	出席	出席	欠席	欠席	欠席	出席	出席	出席	欠席	47
00324	女性	出席	出席	出席	出席	出席	欠席	欠席	出席	欠席	欠席	出席	出席	欠席	出席	出席	62
00367	女性	出席	出席	出席	欠席	出席	出席	出席	欠席	出席	出席	欠席	欠席	出席	欠席	出席	75
00437	女性	出席	欠席	出席	欠席	出席	出席	欠席	欠席	出席	出席	欠席	出席	出席	出席	出席	55
00479	男性	出席	出席	欠席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	95
00522	男性	出席	出席	出席	出席	出席	欠席	出席	欠席	欠席	出席	出席	欠席	欠席	欠席	出席	47
00526	男性	欠席	欠席	出席	出席	欠席	出席	出席	出席	出席	出席	出席	出席	出席	出席	欠席	83
00588	女性	出席	出席	出席	出席	出席	欠席	出席	出席	出席	欠席	欠席	出席	出席	出席	出席	64
00663	男性	出席	出席	出席	欠席	欠席	出席	出席	出席	欠席	出席	出席	欠席	出席	出席	出席	81
00822	女性	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	欠席	出席	40
00830	男性	出席	出席	欠席	出席	欠席	出席	欠席	出席	出席	出席	出席	欠席	出席	出席	出席	66
00855	女性	欠席	欠席	出席	欠席	欠席	出席	出席	出席	出席	欠席	出席	出席	出席	出席	出席	76
00871	女性	出席	出席	出席	出席	出席	出席	欠席	欠席	欠席	出席	出席	出席	出席	出席	出席	43
00977	女性	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	欠席	出席	出席	出席	出席	94
00990	女性	欠席	出席	出席	出席	出席	出席	出席	出席	出席	出席	欠席	欠席	欠席	出席	出席	45
01026	女性	出席	出席	欠席	出席	出席	出席	欠席	出席	欠席	欠席	出席	出席	出席	出席	出席	59
01028	女性	出席	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	出席	80
01035	男性	欠席	欠席	出席	出席	出席	出席	出席	出席	出席	欠席	出席	出席	出席	出席	出席	89
01053	女性	欠席	出席	出席	欠席	出席	出席	出席	出席	欠席	出席	出席	欠席	出席	欠席	欠席	37
01142	女性	出席	出席	出席	出席	出席	出席	出席	出席	欠席	出席	出席	出席	出席	出席	出席	98
01184	男性	出席	出席	欠席	出席	出席	出席	出席	出席	欠席	出席	出席	出席	出席	出席	出席	93
01293	男性	欠席	出席	出席	欠席	出席	欠席	出席	出席	出席	欠席	出席	出席	出席	出席	欠席	79
01470	男性	出席	欠席	出席	出席	出席	出席	出席	欠席	出席	欠席	欠席	出席	出席	出席	出席	83

■ ID = 学生番号

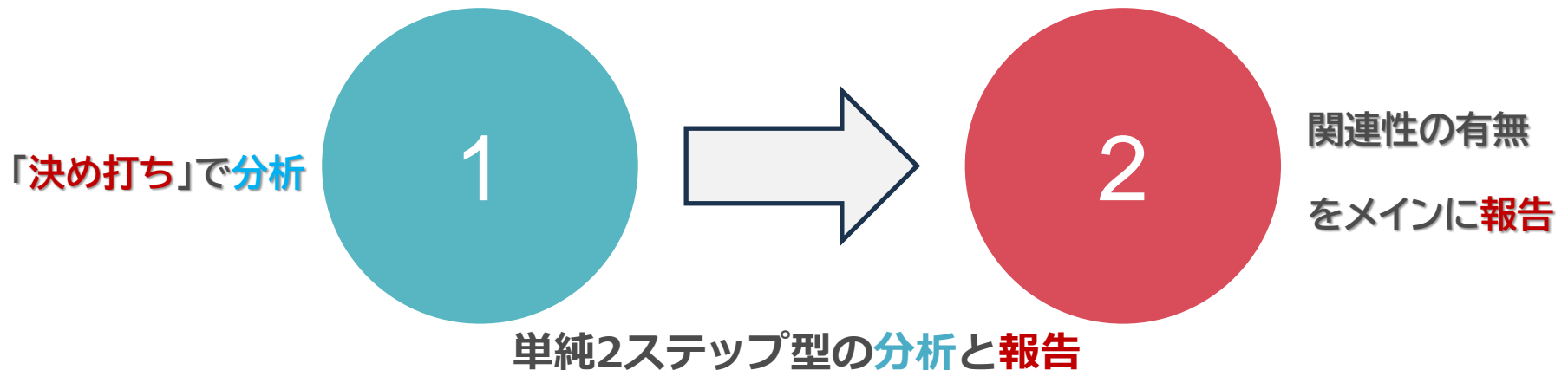
C01 = とある科目における初回の出欠状況

C15 = 最終回の出欠状況

成績素点は100点満点

このような「分析 → 報告」をしていますか？

分析例: 「出席回数」と「成績素点」の関連性を調べて報告



■ みなさんに考えてほしいこと

- 前ページのデータから得られる示唆は「これだけ」ですか？
- 他にも分析できそうな問い（RQ）はありませんか？
- 「このデータの特徴を理解した」と副学長に言えますか？

探索的データ分析（EDA）とは

■ Exploratory Data Analysis (EDA)

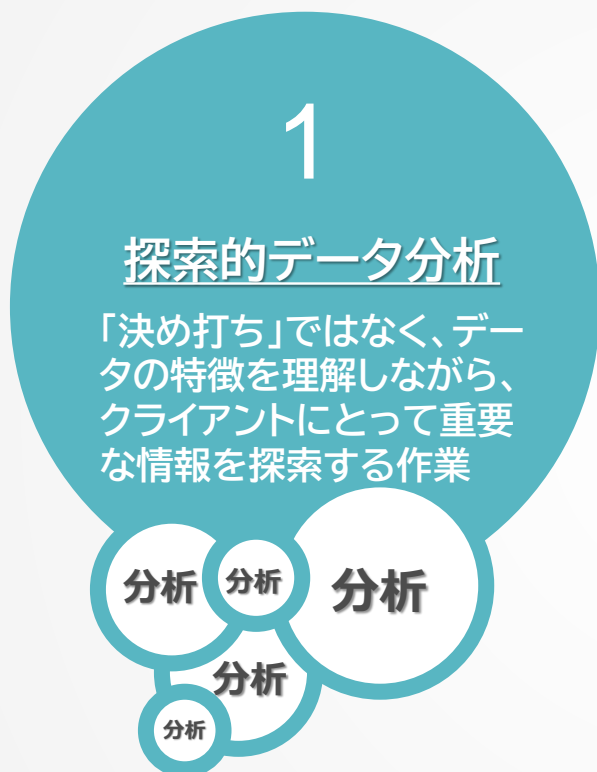
- データ分析における**最初のステップ**
- **【報告する前】** 様々な視点からデータを**集計・可視化**し、**データの特徴を理解**する過程のこと
- データに関する**エラー**を見つけたり、**何らかのパターン・傾向**を見出したりするプロセス

→ **リサーチクエスチョン（RQ）の策定と深い関係がある**

→ **The better you know your data, the better is your analysis.**

データ分析の教科書に出てくる決まり文句

このような「データ分析」を心がけてみよう！



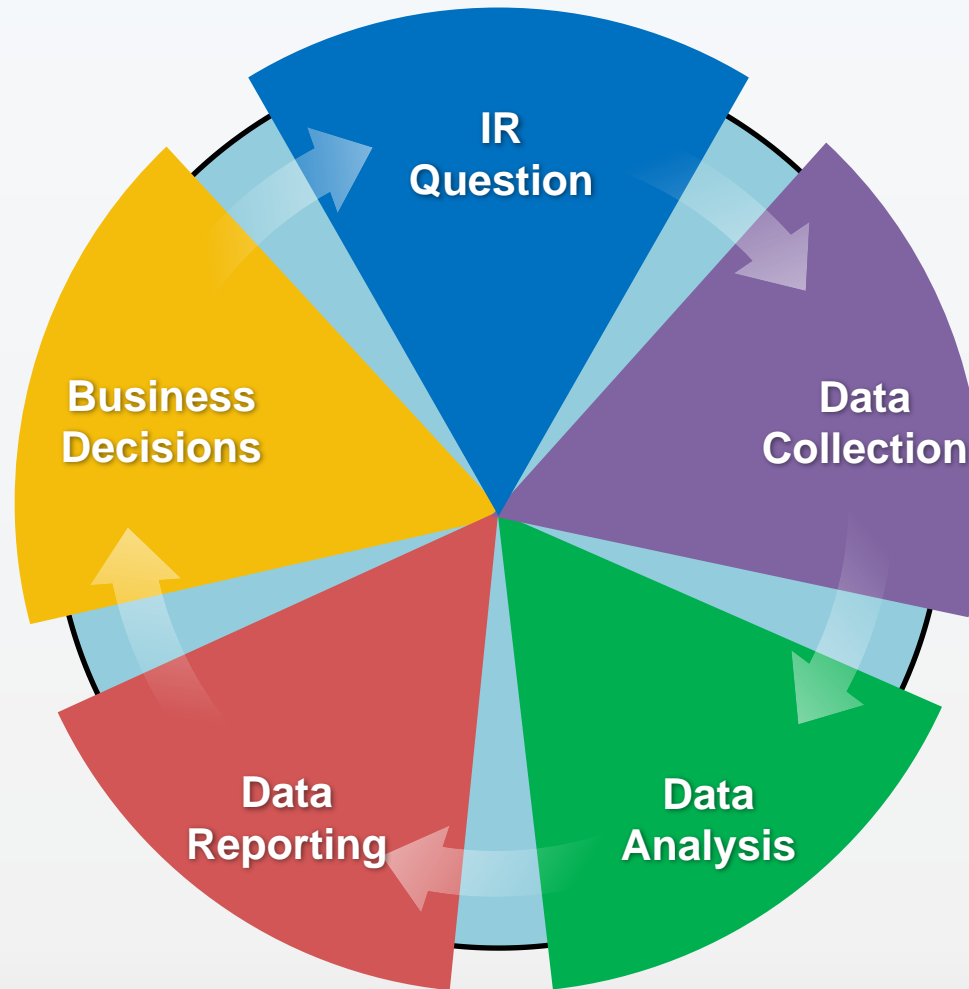
EDAとRQ

- データの特徴をより良く理解するためには、様々な視点からデータを集計・可視化する必要があります。
→ **たくさんの分析作業が必要です。**
- EDAを通じて、IR担当者の問いやRQは、より洗練された物になります。
→ **分析結果が、次の新しい問いやRQに繋がり、データを深く理解できるようになります。**

EDAの注意点

- EDAを起点とした「Data-Drivenな分析」は、慣れるまでは「問い」や「RQ」を立てること自体が困難です。
 - 「経験を積むこと & 練習」が必要
- データの特徴を深く理解するためには、効率的なデータ処理・分析等のスキル※が必須となります。（たくさんの分析が必要となるので）
- EDAは宝探しに似ています。全ての分析結果が報告する価値のあるものとは限りません。ですので、手作業からの脱却を考えましょう。（楽しんで分析！）


意思決定プロセスにおけるIRの役割



Quality in, Quality out.

とある科目の出欠と成績：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」

■ 少し複雑な「問い」の例 ■

午後の演習問題

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」

■ 「問い」は他にもたくさんあります。面白そうな「問い」を考えてみましょう。
この作業が、みなさんの経験値を上げることに繋がります。

【午後の演習関連】 3回以上連続欠席とは？

「p.8」のデータ例より抜粋 + データ追加 (ID = XXXXX)

ID	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
00136	欠席	欠席	欠席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席	出席
00236	出席	出席	欠席	欠席	出席	出席	出席	出席	欠席	欠席	欠席	出席	出席	出席	出席
00522	出席	出席	出席	出席	出席	欠席	出席	欠席	欠席	出席	出席	欠席	欠席	欠席	出席
XXXXX	出席	欠席	欠席	欠席	欠席	欠席	欠席	欠席	欠席	欠席	欠席	欠席	欠席	欠席	出席



ID	3回以上連続欠席とした理由	連続欠席複数回の有無
00136	初回から3回目まで3回連続欠席	X
00236	9回目から11回目まで3回連続欠席	○
00522	12回目から14回目まで3回連続欠席	○
XXXXX	2回目から14回目まで13回連続欠席	X

■ ポイント

- ・ IDごとに、**欠席が3回以上連続しているか**を確認しましょう！

【午後の演習関連】 連続欠席複数回とは？

「p.8」のデータ例より抜粋

ID	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
00067	欠席	出席	欠席	欠席	出席	欠席	欠席	出席	出席	欠席	出席	出席	欠席	出席	欠席
00161	欠席	欠席	出席	欠席	欠席	出席	出席	出席	出席	出席	出席	出席	出席	欠席	欠席
00171	出席	欠席	出席	出席	欠席	欠席	出席	欠席	出席	欠席	出席	出席	出席	欠席	出席
00236	出席	出席	欠席	欠席	出席	出席	出席	出席	欠席	欠席	欠席	出席	出席	出席	欠席
00522	出席	出席	出席	出席	出席	欠席	出席	欠席	欠席	出席	出席	欠席	欠席	欠席	出席



ID	連続欠席複数回とした理由	3回以上連続欠席の有無
00067	3回目と4回目で連続欠席；6回目と7回目で連続欠席	X
00161	初回と2回目で連続欠席；4回目と5回目で連続欠席；14回目と最終回で連続欠席	X
00171	5回目と6回目で連続欠席；13回目と14回目で連続欠席	X
00236	3回目と4回目で連続欠席；9回目から11回目まで3回連続欠席	○
00522	8回目と9回目で連続欠席；12回目から14回目まで3回連続欠席	○

■ ポイント

- 連続欠席が「複数回ある」ためには、連続欠席と連続欠席の間に「出席」が存在する必要があります。
→ 前ページの「ID = XXXXX」を確認

午後のデータ演習について

13:00 – 14:30



演習の目的：「本セッションの案内※」より抜粋

- 前半は講義、後半は「**IRデータ処理の体験演習**」…
- **演習では、パソコンを使いますので「必ず」**
ご持参ください。
- **エクセルの操作方法等に関しては、**
その詳細を扱いません。（時間が足りません。）

ポイント

- ✓ 演習の目的は、
データ分析のリアルを体験することです。
- ✓ 正しい答えを出すことではありません。（それは別の機会で）
- ✓ 間違えてもOK。**仲間と気負わず楽しくやりましょう！**
（ネットワーク作りにも役立ててください。）

演習のシナリオ

- 教学担当副学長から、
「とある科目の出席データと科目成績」を渡され、
以下の緊急分析依頼がありました。

【依頼条件】

約30分以内に、少なくとも「問1」の回答を提出すること

1. 【必須問題】

「3回以上連続欠席」のある学生の
「人数」と「科目平均点数」を「性別ごと」に知りたい。

2. 【余力があれば】

「連続欠席が複数回」ある学生の
「人数」と「科目平均点数」を「性別ごと」に知りたい。

演習用データの場所 (1)

A) 「Google」等で「OIRE」をキーワード検索



クリック

B) HP画面上部の「活動内容」をクリック

C) URL欄にある

「<https://ir.yamagata-u.ac.jp/activity/>」

の後に、**20230831** と入力

演習用データの場所 (2)

OIRE

OIREについて IR担当者向け実践プログラム 活動内容 発表資料 山形大学 Yamagata University

IR初級者セッション

2023.8.31

Handout

・ [演習用データ \(エクセルファイル\)](#) ← クリック

■ 演習用データについて

- ・ このデータは、発表者が**ランダムに生成**した物なので、ご自由にご利用ください。出典も必要ありません。

演習の進め方

■ グループワーク形式

- ・ 各班にファシリテーターが付きます。

■ 進行の目安

1. 自己紹介	5分
2. 個人作業 & グループ会議	20分
3. グループワーク (分担作業をススメます)	25分
4. 結果発表 (各班)	10分

分析依頼への
対応時間

→ 上記は、あくまでも「目安」です。
ファシリテーターの進行に従ってください。

■ 結果発表の方法について

- ・ 14時までに結果を藤原に教えてください。

データ分析の方法について

- 自由に取り組んでください！
 - ただし、今回の依頼条件である「30分以内に、少なくとも問1の回答を提出すること」は、忘れないように！ → 「分担して手作業」も立派な作戦
- 生成系AIの利用も可とします。
 - 今回のデータは「演習用データ」なので、ネットに上げて構いませんが、実践では、所属大学のルールを守りましょう。
- グループワーク終了後、解説編※があります。
 1. ChatGPTを使った分析例 (by 田中秀典)
 2. EDAを効率化するための汎用データについて (by 藤原)
 3. エクセルによる分析例 & 演習まとめ：IR 担当者としてのNext Step (by 山本幸一)

※ 解説編における「資料の場所」は、田中先生から指示があります。

まとめとして



最後に考えてほしいこと（1）

- 今回は、とある**一科目**についての**データ**を扱いました。
- しかし、実際、学生は**一学期に10科目以上履修**します。
 - 大学全体では開講科目が数千以上になることも
- このことは、処理しなければならない**データが膨大**であることを意味します。
- IR担当者として、今後、どのようなスキルが必要となっていくか、考えてみましょう。

最後に考えてほしいこと (2)

- 「決め打ちの分析」は、「適切なRQを立てた分析」ではありません。
- 「クライアントに役立つ情報」を見つけるには、「EDA」が必要です。 (が、見つかるかは分かりません)
- IR担当者にとって重要な役割は意思決定者に「必要と思われる」情報を届けること

∴ 「**Quality in, Quality out.**※」

の考え方 (p.13) が重要だと思います。

The better
you know your data,
the better
is your analysis.

※ 「Garbage In, Garbage Out.」の方が有名かもしれません。

【おまけ】 EDAに関する捉え方の違い

- シナリオ：20の視点から分析を行い、そのうち「2つの結果」を報告

加点主義的な考え方

- 20の気づきを得た！
- 分析スキルと経験値を得た！
- データを深く知ることができた！
- 今回の報告には使えなくても、何処かの場面で、今回の結果が使える可能性がある！
- 20の分析を行ったこと「も」評価の対象

減点主義的な考え方

- 18の分析にかけた時間が無駄
- 無駄な分析 = 効率的な業務ができていない
- 必要なのは「報告内容」であって、「データを知る」ことではない
- 今回の報告に使えるかが重要であって、プロセスは関係ない。次の分析は、他の人がやる可能性が高い。
- $20 - 2 = 「18」$ の分析は評価の対象外

THANK YOU!

ANY QUESTIONS, COMMENTS OR SUGGESTIONS?

藤原 宏司 | Koji Fujiwara, Ph.D.

kfujiwara@cc.yamagata-u.ac.jp



グループワークの結果比較



【グループワーク】 結果比較 : 問 1

【必須】 3 回以上連続欠席

男性

グループ	人数	平均点
#1 田中	29	40.34
#2 紺田	29	40.03
#3 岡部	26	41.80
#4 橋本	29	40.03
#5 山本	29	40.03
#6 小湊	29	40.03
#7 江端	29	40.03

女性

グループ	人数	平均点
#1 田中	35	40.03
#2 紺田	36	40.14
#3 岡部	33	41.70
#4 橋本	36	40.13
#5 山本	36	40.14
#6 小湊	36	40.13
#7 江端	36	40.13

【グループワーク】 結果比較：問2

■ 連続欠席が複数回

男性

グループ	人数	平均点
#1 田中		
#2 紺田		
#3 岡部		
#4 橋本		
#5 山本	27	44.22
#6 小湊	27	44.22
#7 江端		

女性

グループ	人数	平均点
#1 田中		
#2 紺田		
#3 岡部		
#4 橋本		
#5 山本	37	45.03
#6 小湊	37	45.03
#7 江端		