

令和5年11月20日 数理・データサイエンス・AI教育
強化拠点コンソーシアム
第2回コンソーシアム中国ブロックシンポジウム



認定期間 2026.3.31



認定期間 2027.3.31

数理・データサイエンス・AI
教育教科拠点コンソーシアム

大学リーグやまぐちデータサイエンスWGの 取り組みのご紹介



YAMAGUCHI UNIVERSITY
山口大学

教育・学生支援機構
情報・データ科学教育センター

山口大学情報・データ科学教育センター長
山口 真悟

1. 大学リーグやまぐち

1-1 大学リーグやまぐち

2016（H28）年10月、山口県内の高等教育機関の連携を深め、また、行政、産業界等と広範なネットワークを形成し、それぞれの特性を活かした様々な連携事業の実施を通じて、**若者の定着促進並びに高等教育機関の地域貢献力及び教育・研究水準の一層の向上**を図ることにより、地域社会の発展に寄与することを目的に設立*1。



平成28年10月12日 設立会議



*1 <https://www.pref.yamaguchi.lg.jp/soshiki/3/11978.html>

1-2 大学リーグやまぐち データサイエンスWG

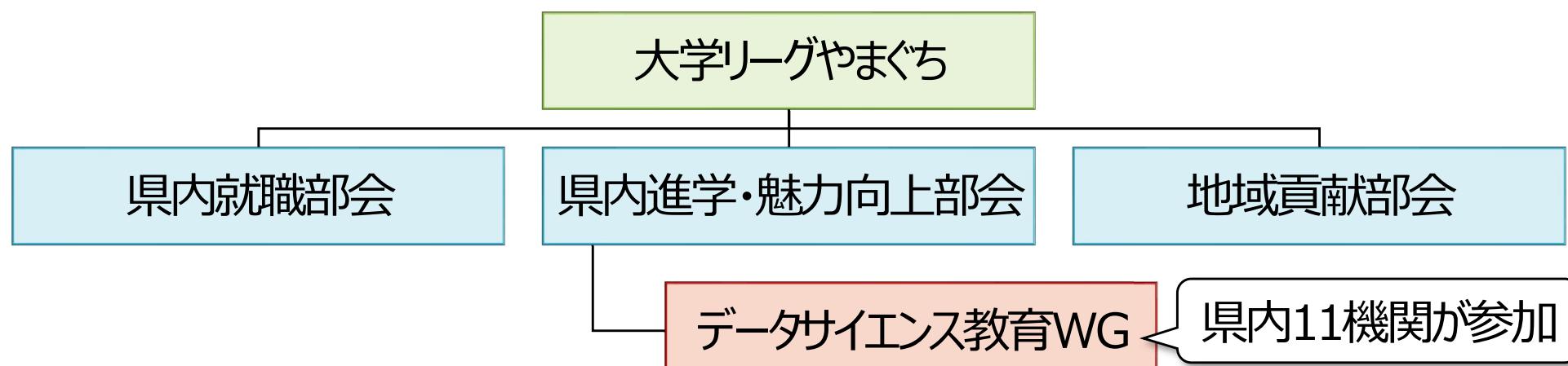
【目的】

データサイエンス教育に関する情報共有等を通じて、県内高等教育機関におけるデータサイエンス教育の普及と促進を図る。2021（R3）年度に設置。

【所掌事項】

- ・ データサイエンス教育に係る情報交換及び課題解決に関する事項
- ・ データサイエンス教育に係る連携に関する事項
- ・ その他データサイエンス教育に関して必要と認められる事項

2022（R4）年度から共通教材の作成に取り組んでいる



1-3 参加校

県内の11機関

国立大学	<u>山口大</u>			
公立大学	<u>山口県立大</u>	<u>周南公立大学</u>	<u>山陽小野田市立 山口東京理科大</u>	下関市立大
私立大学	至誠館大	宇部フロンティア大 同大短期大学部	山口学芸大 山口芸術短期大	
短期大学	<u>山口短期大</u>			
高専	<u>宇部工業高専</u>	<u>徳山工業高専</u>		

※青色下線はリテラシーレベル認定校、赤色下線は応用基礎レベル認定校
 ※令和5年度、山口県立大と宇部工業高専、徳山工業高専がリテラシー
 レベルに新たに認定された

1-4 令和5年度のWGの開催

2023 (R05) 年8月29日 第1回

- ①自己紹介・各校のデータサイエンス教育の取組状況について
- ②山口県警察本部のビデオ教材の紹介
- ③DS共通教育教材作成における著作権の勉強会
山口大学知的財産センター長 小川 明子 教授
- ④共通教育教材サイトの運用について



勉強会テキスト*1

同日午後 数理・データサイエンス・AI教育強化拠点コンソーシアム中国
ブロック拠点校の広島大学との意見交換会

2023 (R05) 年12月予定 第2回

2024 (R06) 年3月予定 第3回

*1 小川明子, 楽しい著作権法, <https://www.tlo.sangaku.yamaguchi-u.ac.jp/liblary/>

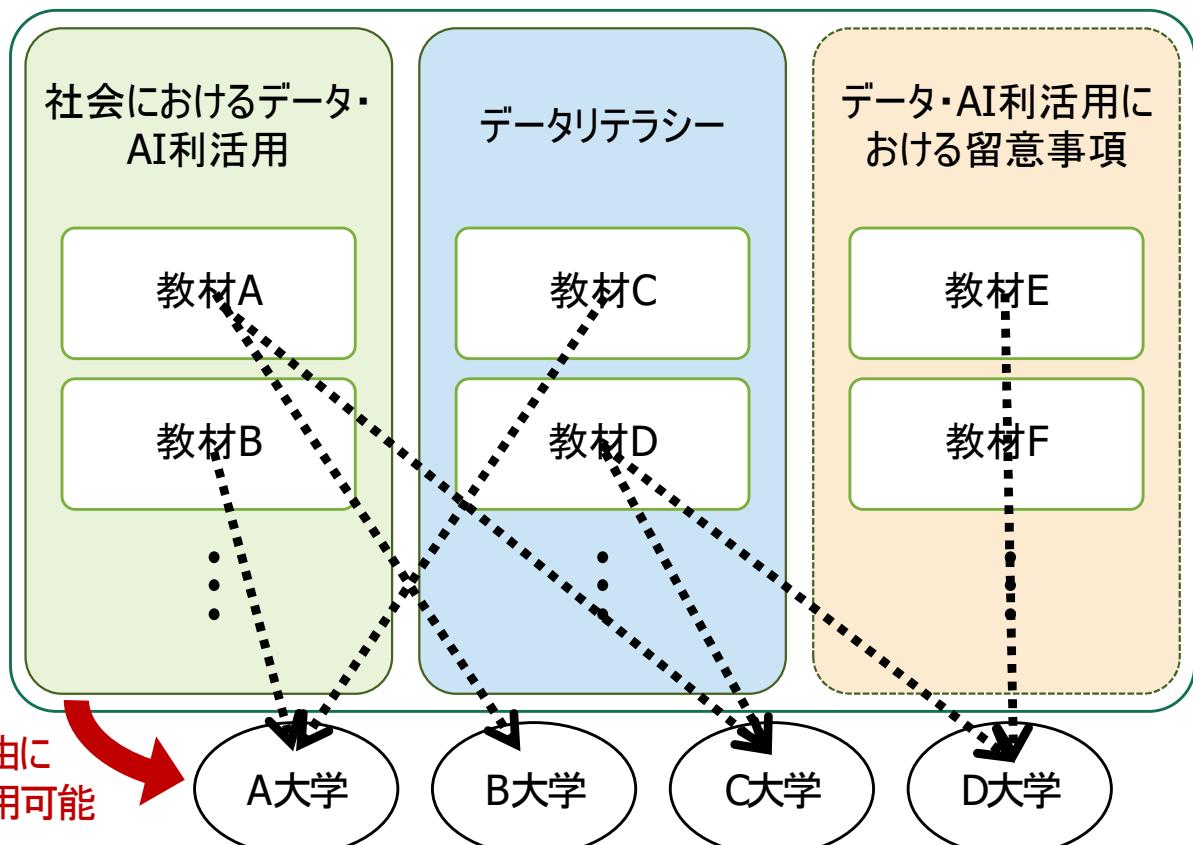
2. 共通教材

2-1 作成方針

山口県内の高等教育機関で実施するデータサイエンス教育において、各機関が**自由に利用可能な山口県らしさ**のある教材を作成する。

【教材レベル】

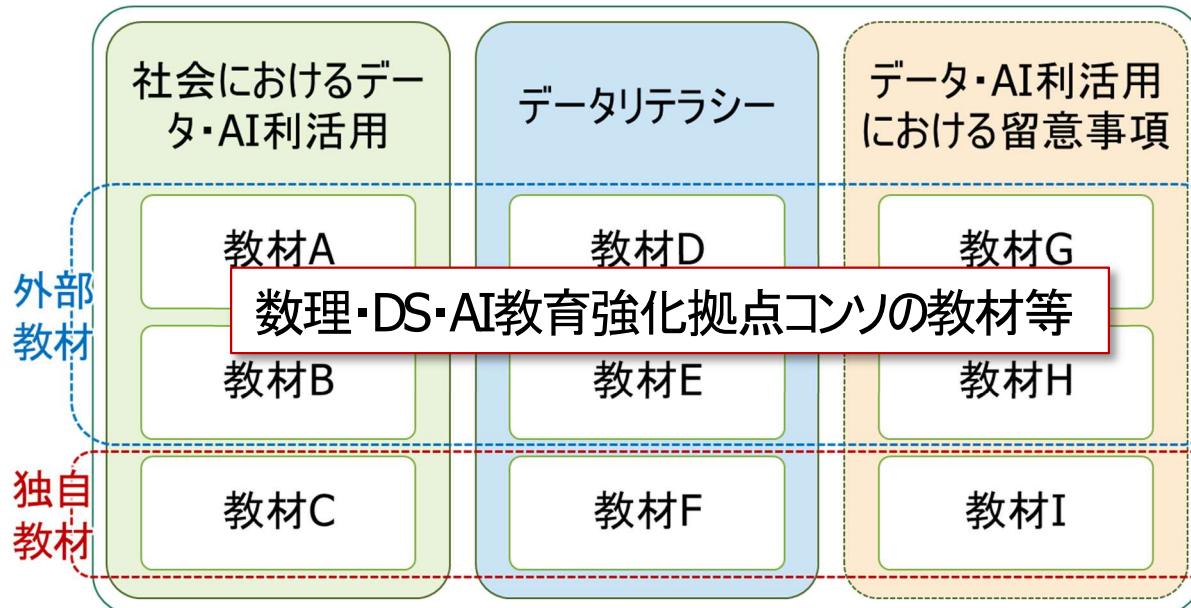
数理・データサイエンス・AI
教育拠点コンソーシアムの
定めるモデルカリキュラム
(リテラシーレベル) に準拠



2-2 教材構成

- ・自前主義に拘らない。既存教材を積極的に活用し、必要なところだけ作る

大学リーグやまぐち共通教材 = 外部教材 + 独自教材



- ・独自教材の例

- ・J2レノファ山口の勝因分析を題材としたデータリテラシー教材
- ・県警と共同開発した学生に身近なセキュリティ教材

2-3 共通教材プロトタイプ（上部）

<https://ds0n.cc>

大学リーグやまぐち・データサイエンス教育ワーキンググループ 共通教育教材（プロトタイプ）

本教材は山口県内でのデータサイエンス教育の普及及び展開促進のため、大学リーグやまぐち・データサイエンス教育ワーキンググループが共同開発したものです。

特徴

- 数理・データサイエンス・AI教育拠点コンソーシアムの定めるモデルカリキュラム（リテラシーレベル）に準拠しています。
- 「大学リーグやまぐち独自教材」には山口県内のデータや事例を盛り込んだ内容を含みます。
- デジタルを基本としつつ、紙印刷にも対応可能な内容・構成です。

利用条件

本教材は

- 数理・データサイエンス・AI教育拠点コンソーシアムが提供する「コンソーシアム教材」
- 大学リーグやまぐち・データサイエンス教育ワーキンググループが提供する「大学リーグやまぐち独自教材」

からなります。

- 「コンソーシアム教材」の利用条件は[こちら](#)をご参照ください
 - 「大学リーグやまぐち独自教材」は、
 - 著作権者を表示すること
 - 山口県内の高等教育機関における非営利目的での利用に限定すること
 - いかなる改変も禁止すること（教材の全部または一部をそのまま利用すること意味する）
- を条件に利用する（再配布や教材の部分的な利用を含む）ことができます。

教材

- [東京大学]は東京大学 数理・情報教育センターが開発した教材です。
- [山口大学]は山口大学 情報・データ科学教育センターが中心に開発した教材です。

1. 社会におけるデータ・AI利活用

利用条件
県内の高等教育機関は
自由に利用できるように
定めています

2-4 共通教材プロトタイプ（下部）

2. データリテラシー

節	コンソーシアム教材	大学リーグやまぐち独自教材
2-1. データを読む	[東京大学]	
2-2. 外部教材	[東京大学]	[山口大学「J2レノファ山口の勝因分析を題材としたデータリテラシー教材」 PPT,PDF]
2-3. 留意事項	[東京大学]	

独自教材

節	コンソーシアム教材	大学リーグやまぐち独自教材
3-1. データ・AIを扱う上での留意事項	[東京大学]	[山口大学「サイバー犯罪の現状と対策（山口県警）」 MP4,PDF]
3-2. データを守る上での留意事項		

4. オプション

節	コンソーシアム教材	大学リーグやまぐち独自教材
4-1. 統計および数理基礎	[東京大学]	
4-2. アルゴリズム基礎	[東京大学]	
4-3. データ構造とプログラミング基礎	[東京大学]	
4-4. 時系列データの解析	[東京大学]	
4-5. テキスト解析	[東京大学]	
4-6. 画像解析	[東京大学]	
4-7. データハンドリング	[東京大学]	
4-8. データ活用実践（教師あり学習）	[東京大学]	[山口大学「機械学習、データサイエンスによる問題解決」 PPT,PDF]
4-9. データ活用実践（教師なし学習）	[東京大学]	

利用にあたってのお願い

本教材は皆様のご意見を反映し、隨時改訂・改善を行っていく予定です。誤りや不明な点、改善すべき点などにお気づきの方は、下記事務局までご連絡ください。

[大学リーグやまぐち・データサイエンス教育ワーキンググループ](#)

事務局

山口県学事文書課

Tel : 083-933-2140

E-mail: leagueyamaguchi#pref.yamaguchi.lg.jp

※上記の#をアットマークに置き換えて下さい。

2-5 独自教材の例（レノファの勝因分析教材①）

Graduate School of Sciences and Technology for Innovation, Yamaguchi University
V231109 2048
2023（令和5）年度版

J2レノファ山口の勝因分析を題材としたデータリテラシー教材

YAMAGUCHI UNIVERSITY
山口大学
教育・学生支援機構
情報・データ情報・データ科学教育センター
大学院創成科学研究所
山口 真悟 (shingo@Yamaguchi-u.ac.jp)

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved
1/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

本教材の利用について

本教材は

- 著作権者を表示すること
- 山口県内の高等教育機関における非営利目的での利用に限定すること
- いかなる改変も禁止すること（教材の全部または一部をそのまま利用すること意味する）

を条件に利用する（再配布や教材の部分的な利用を含む）ことができます。

利用条件を明記

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

1. データサイエンスとは

1-1 データサイエンス

数学や統計学、機械学習、プログラミングなどの知識や技術を活用して、社会に溢れているデータから「価値」を引き出す学問。

デジタル革新と人々の想像・創造力をつなげる架け橋となる学問

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved
3/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

1-2 データサイエンスは「目的」でなく「手段」

「データサイエンスをやってみたい」という相談をよく受けようになりました。

しかし、まずお伝えしたいことは

データサイエンスは「目的」ではない

ということです。

データサイエンスは問題解決の「手段」なのです

しかも、自律的に問題を見つけてくれたり、どんな難問でも立ち所に解決してくれるような万能の道具ではありません。

人間が明確な課題を設定し、適切なデータを用いて統計学や機械学習を使うことで、はじめて課題解決の強力な手助けとなるのです。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

4/36

2-6 独自教材の例（レノファの勝因分析教材②）

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

1-3 問題解決のフレームワーク：PPDACサイクル

それでは課題発見から解決まで、どのように進めればよいのでしょうか？

幾つかのやり方がありますが、**PPDACサイクル**がよく使われています。PPDACサイクルはデータを利用した問題解決の定番のフレームワークです。

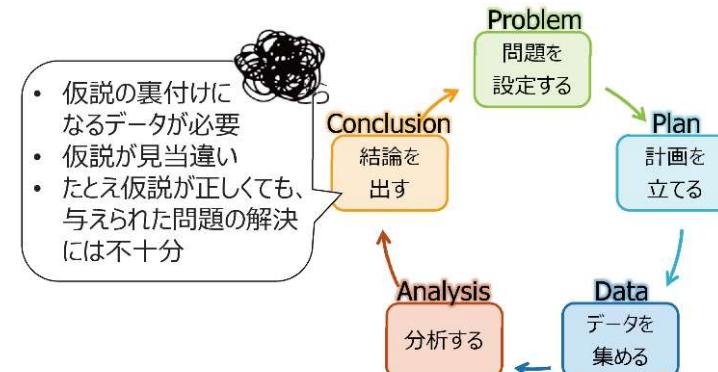
これは、問題解決における各段階を**Problem**（問題設定）、**Plan**（計画）、**Data**（データ収集）、**Analysis**（分析）、**Conclusion**（結論）に分割した考え方です。



Graduate School of Sciences and Technology for Innovation, Yamaguchi University

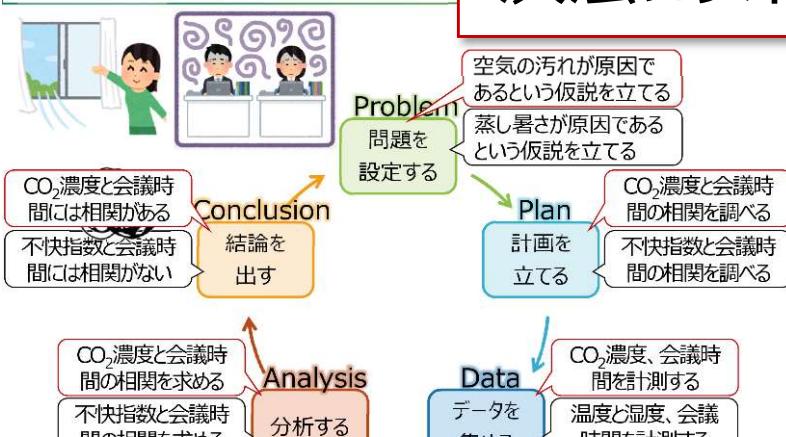
1-4 PPDACサイクル

このプロセスは一度行って終わりではありません。



Graduate School of Sciences and Technology for Innovation, Yamaguchi University

1-5 オフィス環境における生産

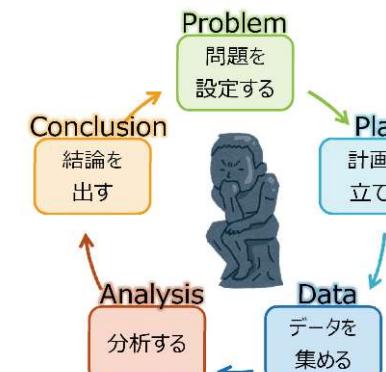


技術より問題解決法にフォーカス

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

1-6 PPDACサイクルこそがデータサイエンスの要

真の問題解決に向けて、サイクルを廻していく。

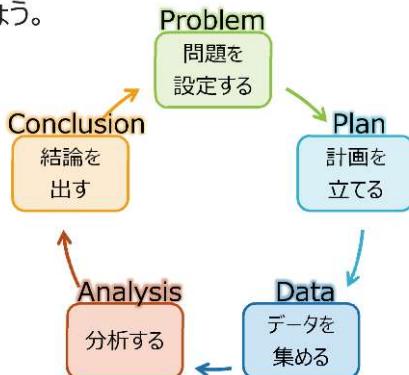


2-7 独自教材の例（レノファの勝因分析教材③）

2. レノファ山口のファンマーケティング

2-1 概要

レノファ山口のファンマーケティングを題材として、データサイエンスの考え方と技術を体験しましょう。



Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

9/36

2-2 レノファ山口FC

2006年創設、山口県全19市町をホームタウンとする日本プロサッカーリーグ（Jリーグ）に加盟するプロサッカークラブです。2016年にJ2リーグへ昇格し、現在J1昇格を目指して奮闘中！



学生に親しみやすい題材

*1 レノファ山口公式ウェブサイト, <https://>

10/36

2-3 レノファ山口の勝因分析

ファンマーケティングの一環として、データサイエンスを活用し、**レノファ山口の勝因**を探ってみましょう。



試合の勝敗にチームの実力が最も大きくかかわることは言うまでもありませんが、それだけではないと考えます。たとえば、

- ・観戦者数
- ・試合会場（ホームかアウェイか）
- ・試合時間帯（デイゲームかナイトゲームか）
- ・天候など、さまざまなもののが考えられます。

レノファ山口の勝因分析を対象にPPDACサイクルを廻していく。

問題解決を一気通貫で体験

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

11/36

2-4 【Problem】問題を設定する

「サッカーJ2リーグのレノファ山口の試合」を対象として、「何が試合を勝利に導くのか」を課題とします。

この課題に対して、評価指標を設定します。評価指標にはさまざまなもののが考えられます。たとえば、



Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

12/36

2-8 独自教材の例（レノファの勝因分析教材④）

2-5 【Problem】勝利に導く要因

要因には以下のようなものが考えられます。

- ・観戦者数
- ・試合会場（ホームかアウェイか）
- ・試合時間帯（デイゲームかナイトゲームか）
- ・天候

さまざま要因があるため、やみくもにいろいろなデータを集めることはなく、できるだけ原因と結果の関係に見通し、仮説を立てて分析を進めることができます。

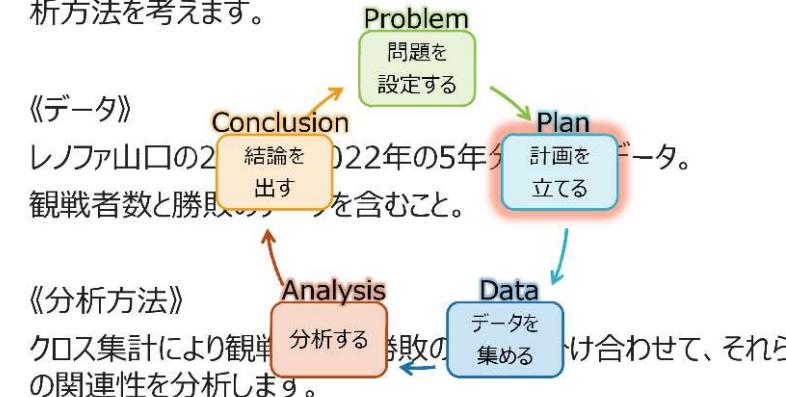
ここでは「観戦者数が勝率に関係する」という仮説を立て、その検証を進めましょう。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

13/36

2-6 【Plan】仮説検証の計画を立てる

仮説「観戦者数が勝率に関する」の検証に必要なデータや分析方法を考えます。



Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

14/36

2-7 【Data】データを収集する

データを実際に収集します。データの取得方法や正確性、信頼性を意識しましょう。

Soccer D.B.*1

*1 Soccer D.B., <https://soccer-db.net/>

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

15/36

2-8 【Data】データの整理

この授業ではMicrosoft Excelを用いてデータを整理します。

試合結果							
	date	home-team	home-score	result	away-score	away-team	attendance
3	2018-02-25	レノファ山口 FC	4	W	1	ロアッソ熊本	7,056
4	2018-03-04	レノファ山口 FC	0	W	0	鹿島FC	1,081
5	2018-03-11	徳島ヴォルティス	2	W	5	レノファ山口 FC	2,081
6	2018-03-17	名古屋トリニティ	3	L	0	レノファ山口 FC	3,401
7	2018-03-21	レノファ山口 FC	2	D	2	サンガ士多	3,601
8	2018-03-25	レノファ山口 FC	0	W	2	柏原FC	4,481
9	2018-03-26	エンタープライズ	0	W	1	レノファ山口 FC	3,600
10	2018-03-31	レノファ山口 FC	2	L	1	大宮アルディージャ	2,116
11	2018-04-07	レノファ山口 FC	2	W	0	レノファ山口 FC	9,676
12	2018-04-14	アビリティ	2	L	0	レノファ山口 FC	3,593
13	2018-04-22	FC今治	1	W	2	レノファ山口 FC	3,935
14	2018-04-28	レノファ山口 FC	1	L	2	ブルビックス柏崎	6,977

データの誤記や未入力、表現の揺れなどの不備を修正し、データの正確性を高めるデータクレンジングという作業が重要です。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

16/36

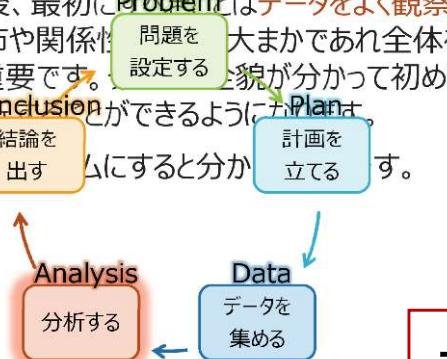
2-9 独自教材の例（レノファの勝因分析教材⑤）

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-9 【Analysis】データの分析

クロス集計により仮説「観戦者数が勝率に関係する」を検証しましょう。データが得られた後、最初に **Problem** は **データをよく観察すること** です。データの分布や関係性を **問題を設定する** 大まかであれ全体を把握することが極めて重要です。**Conclusion** ができるように **Plan**。

データの分布は **Analysis** データを **Data** まとめるにすると分かちます。



Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

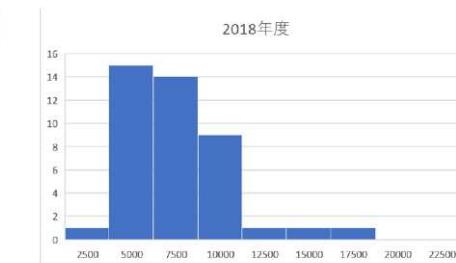
18/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-10 【Analysis】観戦者数の分布

2018年の観戦者数の度数分布表は以下のようになります。階級幅は2,500人としています。ヒストグラムは以下のようにになります。

階級(人)	度数
0~2500	1
2501~5000	15
5001~7500	14
7501~10000	9
10001~12500	1
12501~15000	1
15001~17500	1

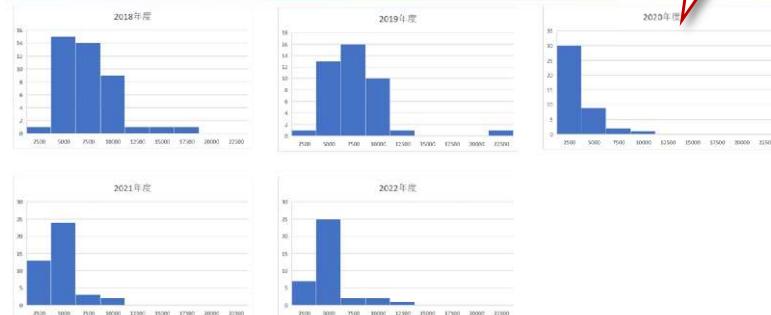


データ全貌の把握の重要性を説く

この分布は単峰性で右裾が長いことが分かります。

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-11 【Analysis】ヒストグラムの考察



- コロナ禍の影響により感染者数の減少がみられます。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

19/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-12 【Analysis】観戦者数-勝敗のクロス集計表

クロス集計とは2種類のデータをかけ合わせて集計し、分析する手法です。

2018年度の観戦者数と勝敗のクロス集計表を作成してみましょう。

表頭

観戦者数(人)	勝敗(試合数)	
	勝ち	引き分け & 負け
0~2500	0	1
	6	9
	7	7
	3	9

表側

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

20/36

2-10 独自教材の例（レノファの勝因分析教材⑥）

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-13 【Analysis】クロス集計表の作り方

- ①クロス集計表のひな形を作成します。

観戦者数(人)	勝ち	引き分け&負け
0～2500	0	1
2501～5000	6	9
5001～7500	7	10
7501～	3	9

```
= COUNTIFS($E$4:$E$45, " = WResult",
           $H$4:$H$45, " >= 0",
           $H$4:$H$45, " <= 2500")
```

- ②勝ちセルに

```
= COUNTIFS($E$4:$E$45, " <> WResult",
           $H$4:$H$45, " >= 0",
           $H$4:$H$45, " <= 2500")
```

引き分け&負けセルに

```
= COUNTIFS($E$4:$E$45, " <> WResult",
           $H$4:$H$45, " > 2500",
           $H$4:$H$45, " <= 5000")
```

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

21/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-14 【Analysis】クロス集計表の作り方

- ③他の行も同様に入力しましょう。

観戦者数(人)	勝ち	引き分け&負け
0～2500	0	1
2501～5000	6	9
5001～7500	7	10
7501～	3	9

```
= COUNTIFS($E$4:$E$45, " = WResult",
           $H$4:$H$45, " > 2500",
           $H$4:$H$45, " <= 5000")
```

引き分け&負けセルに

```
= COUNTIFS($E$4:$E$45, " <> WResult",
           $H$4:$H$45, " > 2500",
           $H$4:$H$45, " <= 5000")
```

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

22/36

プログラミングでなく、エクセルでデータ分析を体験

観戦者数(人)	勝ち	引き分け&負け	勝率
0～2500	0	1	0%
2501～5000	6	9	40%
5001～7500	7	10	50%
7501～	3	9	25%

観戦者数(人)	勝ち	引き分け&負け	勝率
0～2500	0	1	0%
2501～5000	6	5	38%
5001～7500	6	10	38%
7501～	2	10	17%

2020年度

観戦者数(人)	勝ち	引き分け&負け	勝率
0～2500	6	24	20%
2501～5000	1	8	11%
5001～7500	1	1	50%
7501～	1	0	100%

2021年度

観戦者数(人)	勝ち	引き分け&負け	勝率
0～2500	6	7	46%
2501～5000	2	22	8%
5001～7500	2	1	67%
7501～	0	2	0%

2022年度

観戦者数(人)	勝ち	引き分け&負け	勝率
0～2500	4	3	57%
2501～5000	8	19	30%
5001～7500	1	3	25%
7501～	0	4	0%

- 観戦者数と勝率の関連性は年度ごとに異なる結果となりました。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

23/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-16 【Conclusion】分析結果の考察と結論

2018～2022年度

観戦者数(人)	勝ち	引き分け&負け	勝率
0～2500	16	36	31%
2501～5000	22	66	2%
5001～7500	17	22	4%
7501～	6	25	2%

Problem
問題を設定する

5年分の結果をConclusionとも勝率が高かつた。仮説Plan+5001～7500人、次点は0～2500人になった。仮説Plan+5001～7500人難い結果となってしまった。

このように自分が思うような結果が出てこないことは、しばしばあります。

このとき、直ちに仮説Analysisしていると批判されることは危険です。

- データに問題はないか分析する
- 採用した統計手法に問題がないか？
- などを慎重に再検討する必要があります。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

24/36

2-11 独自教材の例（レノファの勝因分析教材⑦）

DX概論第4回ワークシート - Excel

ホーム

貼り付け

クリップボード

フォント

配置

数値

スタイル

セル

編集

A1

1 2018年

index	date	home-team	home-score	result	away-score	away-team	attendance
0	2018/2/25	レノファ山口FC		WResult	1	ロアッソ熊本	7456
1	2018/3/4	レノファ山口FC		WResult	0	愛媛FC	5095
2	2018/3/11	栃木SC		WResult	5	レノファ山口FC	4681
3	2018/3/17	水戸ホーリーホック		LResult	0	レノファ山口FC	4457
4	2018/3/21	レノファ山口FC		DResult	2	ツエーゲン金沢	3508
5	2018/3/25	レノファ山口FC		DResult	2	松本山雅FC	6318
6	2018/3/31	モンテディオ山形		WResult	1	レノファ山口FC	5029
7	2018/4/7	レノファ山口FC		WResult	1	大宮アルディージャ	5353
8	2018/4/14	アビスパ福岡		LResult	0	レノファ山口FC	6837
9	2018/4/22	F C町田ゼルビア		WResult	2	レノファ山口FC	3535
10	2018/4/28	レノファ山口FC		LResult	2	アルビレックス新潟	6374
11	2018/5/3	ヴァンフォーレ甲府		DResult	1	レノファ山口FC	8416
12	2018/5/6	京都サンガF.C.		WResult	2	レノファ山口FC	7093
13	2018/5/12	レノファ山口FC		WResult	3	東京ヴェルディ	4893
14	2018/5/20	大分トリニータ		DResult	2	レノファ山口FC	11862
15	2018/5/27	レノファ山口FC		WResult	0	カマタマーレ讃岐	5034
16	2018/6/2	ジェフユナイテッド千葉		DResult	2	レノファ山口FC	9787
17	2018/6/9	レノファ山口FC		WResult	0	ファジアーノ岡山	7902
18	2018/6/16	徳島ヴォルティス		WResult	2	レノファ山口FC	4828
19	2018/6/23	F C岐阜		DResult	2	レノファ山口FC	8317
20	2018/7/1	レノファ山口FC		LResult	3	横浜FC	12927
21	2018/7/7	ツエーゲン金沢		DResult	2	レノファ山口FC	2108
22	2018/7/15	東京ヴェルディ		LResult	1	レノファ山口FC	5712
23	2018/7/21	レノファ山口FC		DResult	2	水戸ホーリーホック	4909
24	2018/7/25	ロアッソ熊本		DResult	2	レノファ山口FC	3576
25	2018/8/4	愛媛FC					
26	2018/8/12	レノファ山口FC					
27	2018/8/18	レノファ山口FC					

階級

階級	度数
2500	1
5000	15
7500	14
10000	9
12500	1
15000	1
17500	1
20000	0
22500	0
0	0

観戦者数〔人〕

観戦者数〔人〕	勝ち	引き分け＆負け	勝率
0～2500	0	1	0%
2501～5000			#DIV/0!
5001～7500			#DIV/0!
7501～			#DIV/0!

試合会場

試合会場	勝ち	引き分け＆負け	勝率
ホーム	8	13	38%
アウェイ			#DIV/0!

ワークシートを準備

2-12 独自教材の例（レノファの勝因分析教材⑧）

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-17 【Conclusion】問題の解決策を提案する

観戦者について考えると、すべての観戦者がレノファ山口の応援者とは言えず、対戦チームの応援者も含まれます。



特にアウェイの試合では、観戦者の多くは対戦チームの応援者でしょう。

そこで、今度は勝率を高める要因として「試合会場」新たなPPDACサイクルを回すことにしましょう。

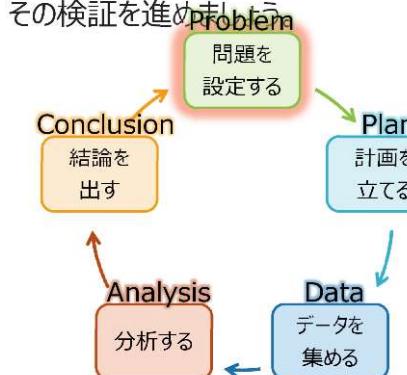
Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

一巡目は失敗 失敗からの学び

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-18 【Problem】勝利に導く要因

「試合会場（ホームかアウェイか）が勝率に関係する」という仮説を立てて、その検証を進めましょう。



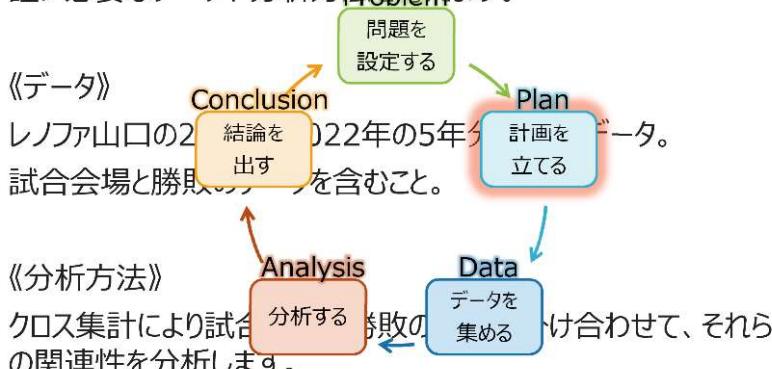
Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

26/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-19 【Plan】仮説検証の計画を立てる

仮説「試合会場（ホームかアウェイか）が勝率に関係する」の検証に必要なデータや分析方法を考えます。



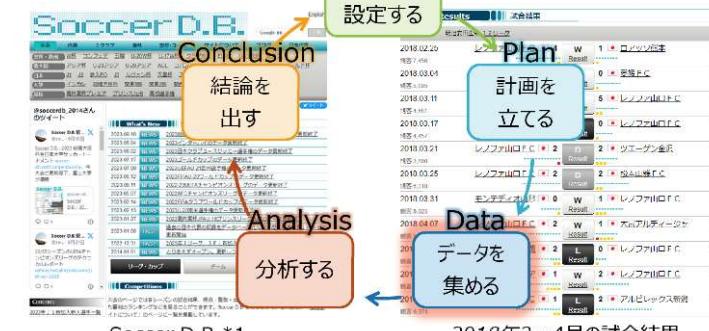
Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

27/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-20 【Data】データを収集する

データを実際に収集します。再び「Soccer D.B.*1」からデータを収集しました。



Soccer D.B.*1

2018年2～4月の試合結果

*1 Soccer D.B., <https://soccer-db.net/>

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

28/36

2-13 独自教材の例（レノファの勝因分析教材⑨）

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-21 【Analysis】データの分析

クロス集計により仮説「試合会場（ホームかアウェイか）が勝率に関係すること」を検証しましょう。Problem

2018年度の試合会場と勝ち分け負けを分析してみましょう。

試合会場	勝ち	引き分け & 負け	勝率
ホーム	8	13	38%
アウェイ	8	13	38%

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-22 【Analysis】クロス集計表の作り方

```
= COUNTIFS($E$4: $E$45, " = WResult", $C$4: $C$45, " = レノファ山口FC")
```

```
= COUNTIFS($E$4: $E$45, " <> WResult", $C$4: $C$45, " = レノファ山口FC")
```

```
= COUNTIFS($E$4: $E$45, " = WResult", $C$4: $C$45, " <> レノファ山口FC")
```

```
= COUNTIFS($E$4: $E$45, " <> WResult", $C$4: $C$45, " <> レノファ山口FC")
```

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved 30/36

二巡目で成功体験

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-23 【Analysis】データの分析

2018年度

試合会場	勝ち	引き分け & 負け	勝率
ホーム	8	13	38%
アウェイ	8	13	38%

2019年度

試合会場	勝ち	引き分け & 負け	勝率
ホーム	8	13	38%
アウェイ	5	16	24%

2020年度

試合会場	勝ち	引き分け & 負け	勝率
ホーム	7	14	33%
アウェイ	2	19	10%

2021年度

試合会場	勝ち	引き分け & 負け	勝率
ホーム	6	15	29%
アウェイ	4	17	19%

2022年度

試合会場	勝ち	引き分け & 負け	勝率
ホーム	8	13	38%
アウェイ	5	16	24%

- 年度に限らず、ホームでの勝率は比較的高い結果となりました。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved 31/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-24 【Conclusion】分析結果の考察と結論

2018～2022年度

試合会場	勝ち	引き分け & 負け	勝率
ホーム	37	68	35%
アウェイ	24	81	23%

Problem

問題を設定する

5年分の結果をみると、ホームでの勝率は比較的高く、アウェイでの勝率は低くなっています。この結果から、「試合会場が勝率に関係する」という仮説が立たなくなっています。仮説の真偽を確認するため、統計的仮説検定をしてみましょう。

クロス集計表の表頭と表側が関連しているのかを検証する方法として「カイ二乗検定」を使われます。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved 32/36

2-14 独自教材の例（レノファの勝因分析教材⑩）

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-25 【Conclusion】統計的仮説検定

有意水準を5%としてカイ二乗検定を使って検証してみましょう。

- ①実測値データを準備します。先ほどのクロス集計表に縦横の合計を追加しましょう。合計は Σ オートSUMを使うと計算できます。

試合会場	勝ち	引き分け&負け	合計
ホーム	37	68	105
アウェイ	24	81	105
合計	61	149	210

- ②期待値データを準備します。

勝敗と試合会場は無関係（独立）であると想定した場合の値

試合会場	勝ち	引き分け & 贠け	合計
ホーム			105
アウェイ			105
合計	61	149	210

カイ二乗検定では実測値と期待値の間に差がない確率（P値）を求めます

- ③CHISQ.TEST関数を使ってP値を計算します。

=CHISQ.TEST(B4:C5,B10:C11) P値 4.82%

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

統計的仮説検定 の体験

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

2-26 【Conclusion】統計的仮説検定

有意水準5%に設定するということは、「5%以下の確率で起こることは、めったに起こらないことであり、何かしら意味がある（有意である）」という意味です。

P値 4.82% \leq 有意水準 5% ですから

「実測値と期待値に差がない確率が低い」ので、
「実測値と期待値の差には何かしら意味がある」と考えます。

以上のことから、仮説「試合会場が勝率に関係する」は5%水準で統計学的に有意であると判断できます。

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

34/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

3. 宿題

「先制点を取ることが勝率に関係する」という仮説を立てて、それを検証して下さい。

- 検証は有意水準1%でカイ二乗検定で行ってください。
- 結果をワークシートに記入して、Moodleを通じて提出して下さい。

仮説を自由に立てさせ、
その立証を実践させる

Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

35/36

Graduate School of Sciences and Technology for Innovation, Yamaguchi University

4. おわりに

4-1 まとめ

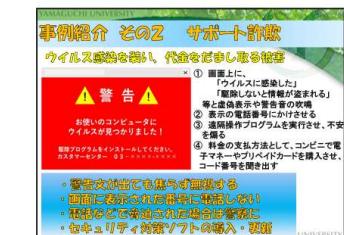
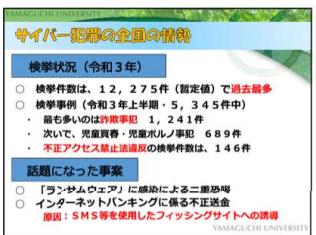
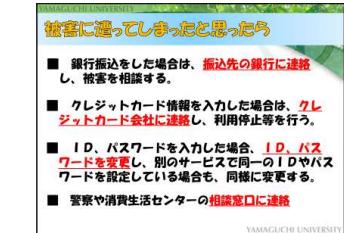
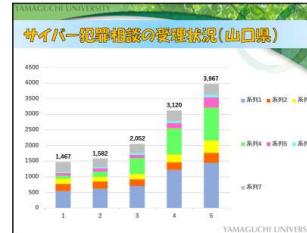
- データサイエンスとは、数学や統計学、機械学習、プログラミングなどの知識や技術を活用して、社会に溢れているデータから「価値」を引き出す学問です。
- データを利用した問題解決の定番のフレームワークとしてPPDACサイクルを紹介しました。
- レノファ山口のファンマーケティングを題材として、データサイエンスの考え方と技術を体験しました。
- 「Think with Data（データと共に思考する）」ことが肝心です。



Copyright (c) 2023 Shingo Yamaguchi, All Rights Reserved

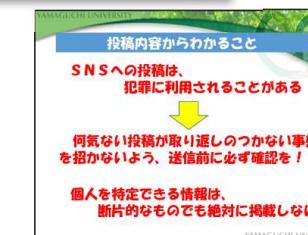
36/36

2-15 独自教材の例（県警によるセキュリティ教材）



県内の最新情報

学生に身近な話題



※上記は令和3年度版。県警の協力により毎年、更新されている。

3. おわりに

3-1 まとめ

- 大学リーグやまぐち データサイエンスWGは、山口県内の高等教育機関におけるデータサイエンス教育の普及と促進を図るため、共通教材を開発しています。
- 共通教材の作成方針を定め、プロトタイプを作成しました。
- 著作権等の取り扱いが依然難しい→専門家と相談し解決していきます。
- プロトタイプの利用を通じて課題や要望を整理し、山口県らしさのあるコンテンツを充実させていきます。

山口大学情報・データ科学教育センター

<https://www.dsc.yamaguchi-u.ac.jp/>
dsm@yamaguchi-u.ac.jp

