

応用基礎レベルの授業向け教材の紹介

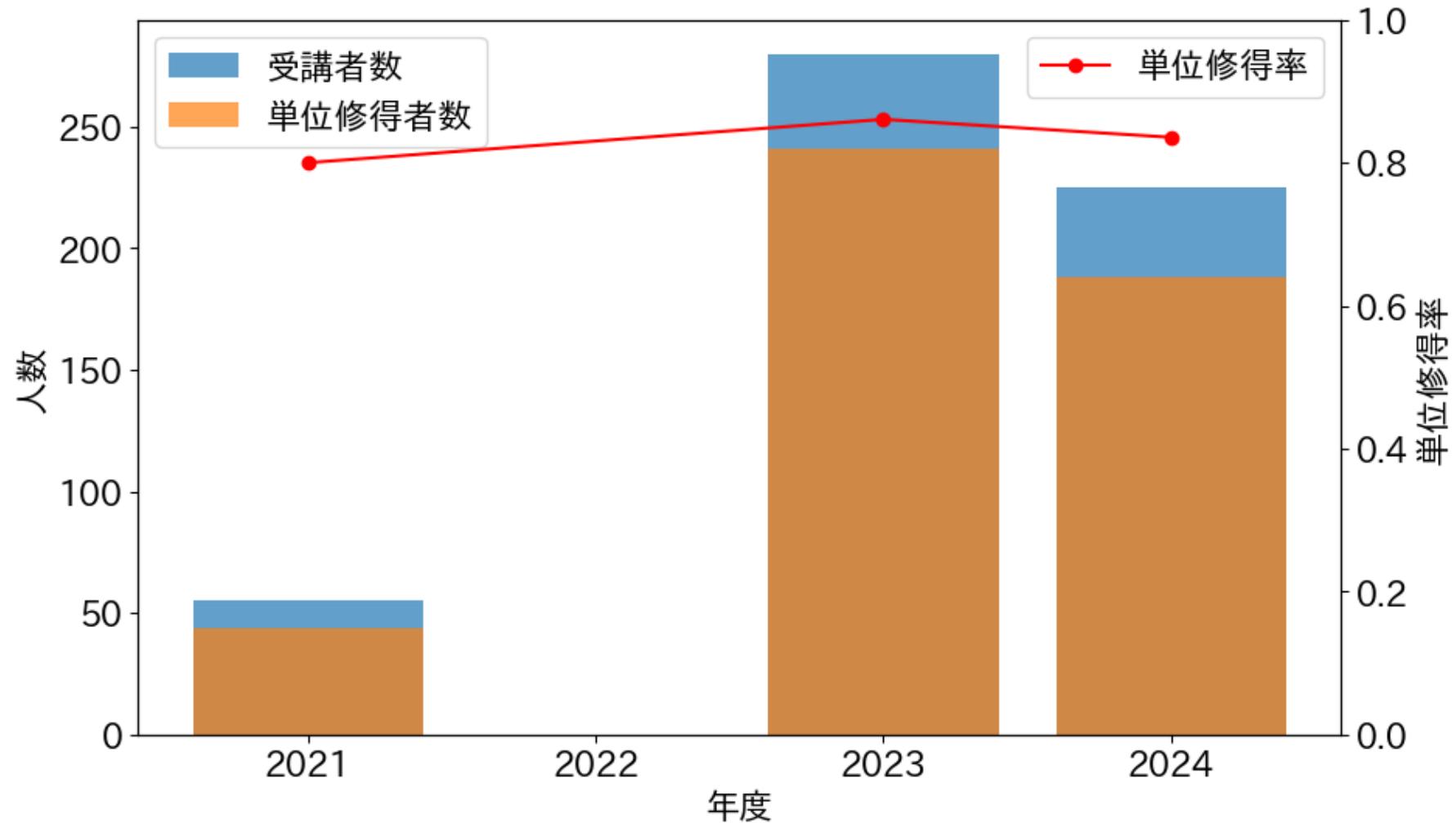


弘前大学 数理・データサイエンス教育センター
増本 広和

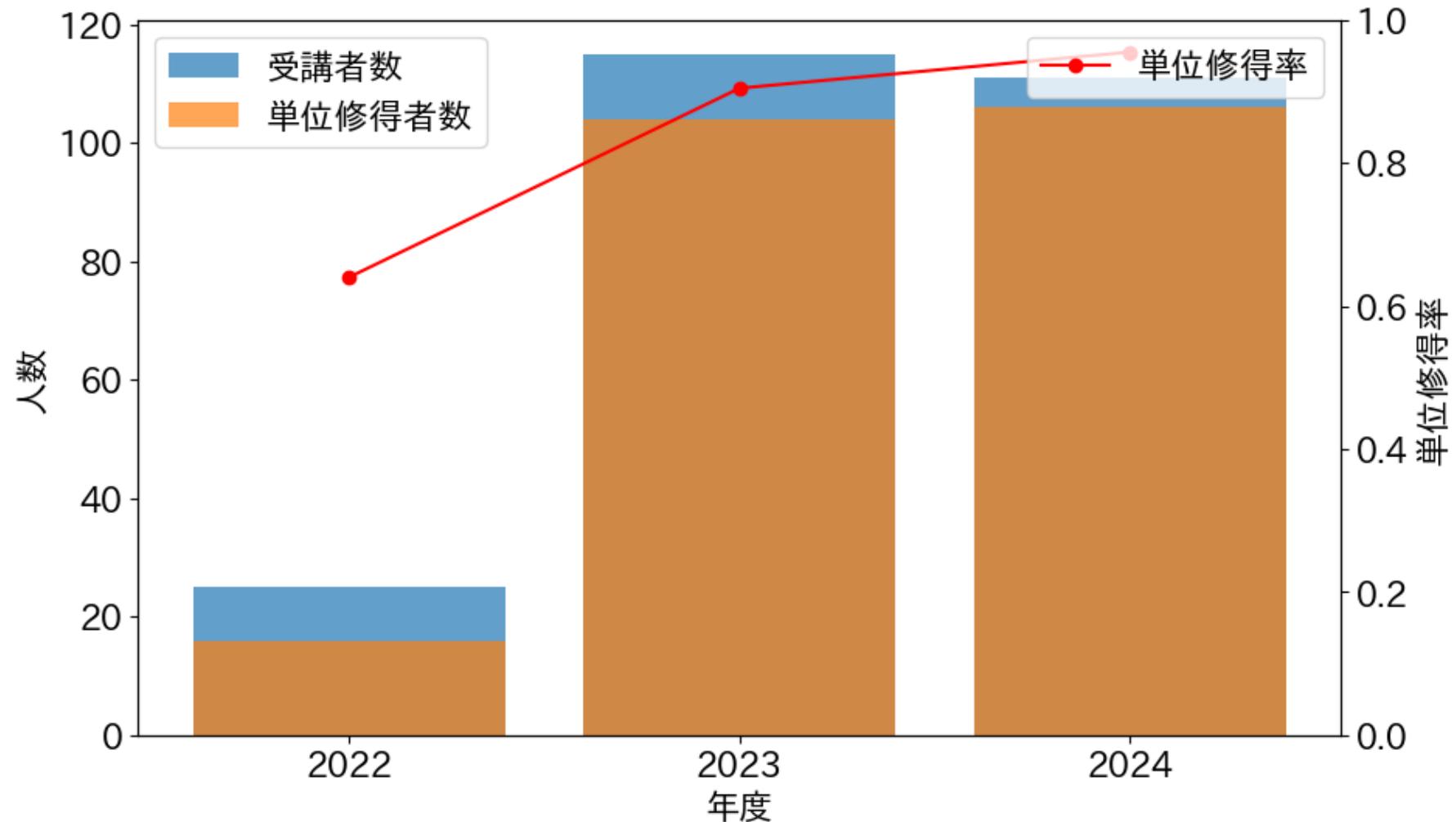
弘前大学のデータサイエンス教育プログラム



データサイエンス発展I 受講者数



データサイエンス発展II 受講者数



データサイエンス発展I 授業の概要

- 「データサイエンス基礎」を履修し、**データ分析について基礎知識のある者**を対象とする
- プログラミング言語Pythonについて基礎から学び、**Pythonを用いたデータ分析**を学ぶ
- **回帰分析**の手法と具体的な課題を用いて、データサイエンスならびに人工知能による実践的な課題解決の方法を学ぶ

データサイエンス発展II 授業の概要

- 「データサイエンス発展I」を履修し、**Pythonプログラミングでデータ分析の経験のある者**を対象とする
- **AIのアルゴリズム**を学び、具体的な課題を用いてデータサイエンスならびに人工知能による実践的な課題解決の方法を学ぶ

DS発展I 各回の内容（令和7年度予定）

1. データサイエンスと人工知能
2. Pythonプログラミングの基礎1: プログラミングとは・変数とデータ型
3. Pythonプログラミングの基礎2: 文字列操作・データ構造
4. Pythonプログラミングの基礎3: リスト操作・論理演算と条件分岐
5. Pythonプログラミングの基礎4: 反復処理・内包表記・関数
6. Pythonプログラミングの基礎5: ライブラリとファイル入出力
7. Pythonプログラミングの基礎（復習）とChatGPTの利用
8. データの特徴を知る1: 確認・操作・データ集計
9. データの特徴を知る2: データの可視化手法
10. 機械学習のためのデータ設計: 汎化能力・交差検証
11. 数値予測問題に取り組む1: データの確認・特徴把握
12. 数値予測問題に取り組む2: 数値予測と回帰モデル
13. 分類問題に取り組む1: ロジスティック回帰
14. 分類問題に取り組む2: 予測モデルの作成
15. 学習状況の確認（試験含む）と解説

データサイエンス発展I 授業の実施について

- 実習室にて対面授業を実施
- 講義とPython言語を用いた実習で構成
- Pythonプログラミング環境: JupyterHub

- 教材: Moodleのコースページに授業スライド, 小テストと実習用Notebookファイルを提示

- 評価: 小テスト, 実習用Notebookとデータ分析課題 (数値予測問題, 分類問題)

- 高大連携公開講座として開講

データサイエンス発展I 教材例



第2回 Pythonプログラミングの基礎1: プログラミングとは・変数とデータ型



講義資料2-1



確認問題2-1

完了 ▾

すべての問題に解答せよ。締め切りは本日中。



プログラミング課題2-1

完了 ▾

以下のファイルの問題に解答し、セーブしたのち、**ipynbファイル**を提出する (締め切り: 次回授業開始時刻)。

※提出ファイルに関する注意

- 提出ファイル形式は、必ずipynbファイルにすること
- 問題セルをすべて実行してからファイルを提出すること
- 問題セルの書式を変えないこと (「#問題1」等を消さない)

以上が満たされていないと正しく採点できません。

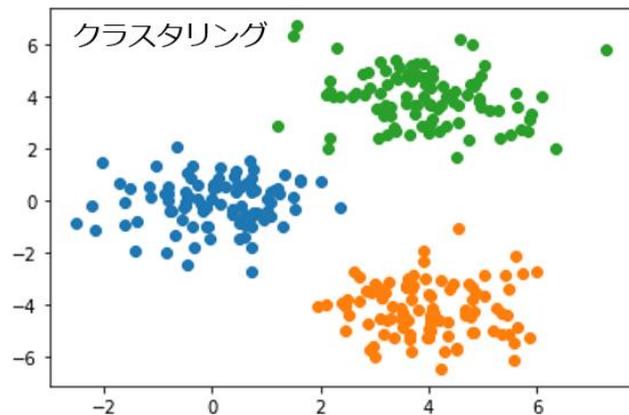


講義資料2-2

データサイエンス発展I 教材例

教師なし学習

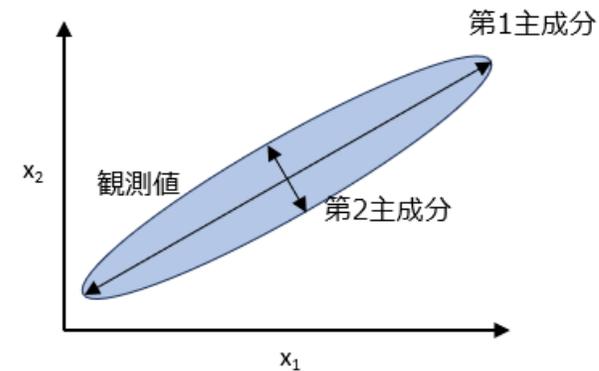
- 入力データのみに基づき、そのデータの特徴をあらわすモデルを学習する
- 次元削減, 異常検出 (機器の故障検知), クラスタリングなど



6

主成分分析

- 主成分分析 (PCA: Principal component analysis) は, 観測値を線形な座標変換によって得られた成分 (主成分) に変換する
- 観測値の分散をある程度維持したまま, より少ない主成分で観測値を表現することが期待される



8

データサイエンス発展I 教材例

問題 1

未解答

最大評点 5.00

▼ 問題にフラグ
を付ける

⚙ 問題を編集
する

v5 (最新)

次の文章のうち正しいものを選択してください。

- 1. コンピュータは、Pythonのプログラムコードを直接理解して処理を行う。
- 2. Pythonのプログラムコードは、大文字と小文字の違いを区別しない。
- 3. Pythonには、1行ずつプログラムコードを入力して、そのつど実行する方法がある。
- 4. 「print(こんにちは)」と記述すると、「こんにちは」という文字列が出力される。
- 5. 変数には後から異なる値を代入できる。

次のページ

データサイエンス発展I 教材例

ランチャー

×

第8回_ライブラリを用いた可視化

+



Markdown



git

Open in...



2. pandasを使ったデータ可視化

pandasには、matplotlibと連携する機能として `plot()` というメソッドが存在します。

`plot()` メソッドを用いることで比較的簡単にDataFrame形式のデータを可視化できます。

pandasでは大まかに下記の種類のグラフの作成が可能です。

ここでは、よく用いられる折れ線グラフ、棒グラフ、散布図、ヒストグラムを紹介します。

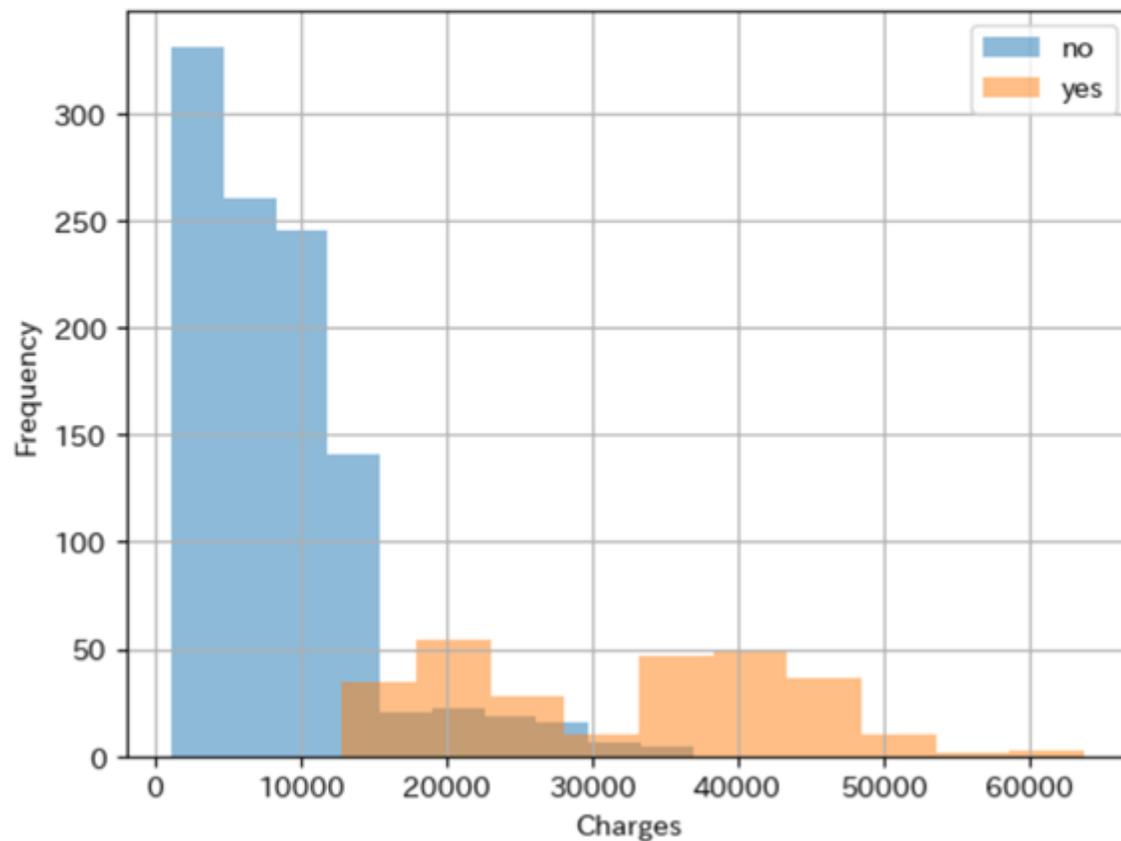
pandasで作成可能なおもなグラフ一覧

グラフの種類	メソッド名
折れ線グラフ	<code>plot()</code>
棒グラフ	<code>plot.bar()</code>
ヒストグラム	<code>plot.hist()</code> または <code>hist()</code>
箱ひげ図	<code>plot.box()</code>
円グラフ	<code>plot.pie()</code>
散布図	<code>plot.scatter()</code>

データサイエンス発展I 教材例

```
ランチャー × 第8回_ライブラリを用いた可視化 × +
+ ✂ 📄 ▶ ■ ↻ ▶▶ Markdown ☰ ⌚ git Open in...

[30]: #問題6
# 喫煙・非喫煙で医療費のヒストグラムをプロット
df.
```



データサイエンス発展I 教材例

ランチャー

×

第10回_回帰モデル_test.ipynb

+



git

Open in...



Python 3 (ipykernel)



1.説明変数がひとつの回帰モデル

Pythonを用いた回帰分析になれるために、説明変数がひとつの回帰モデルから説明します。

データ $(x_i, y_i), i = 1, \dots, n$ に対して、 y を目的変数、 x を説明変数とした回帰モデルは次の式で定義されます。

$$y_i = b_0 + b_1 \cdot x_i + \epsilon_i, i = 1, \dots, n$$

ここで、 ϵ_i は誤差項で、目的変数のうち説明変数の一次式で説明できない部分を表しています。 b_0, b_1 を切片、回帰係数と呼び、誤差項が平均0のとき、次の式で計算することが出来ます。

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

$$b_1 = \frac{S_{x,y}}{S_x^2}.$$

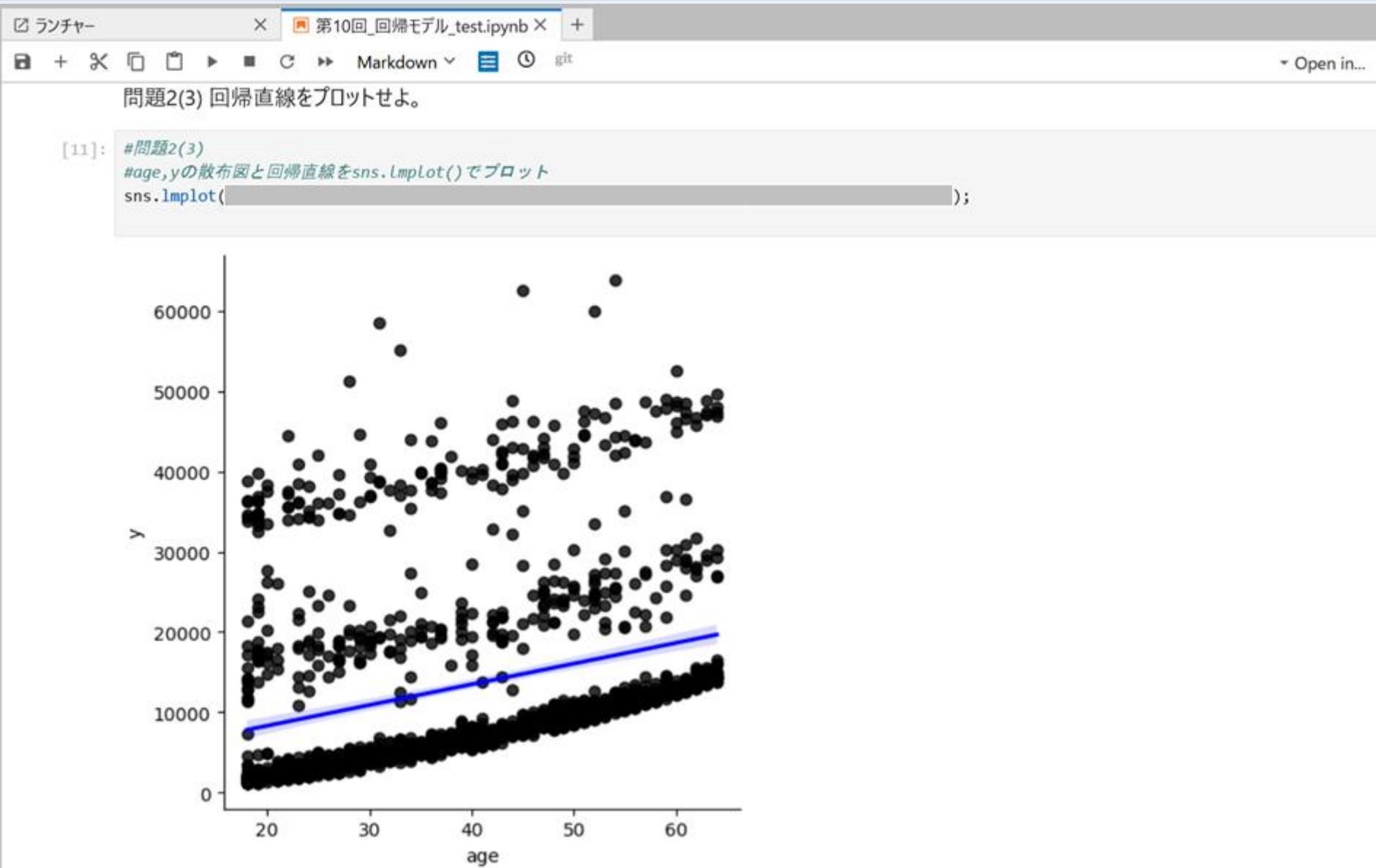
ここで、 \bar{x} は x_i の平均値、 \bar{y} は $y_i, i = 1, \dots, n$ の平均値です。また、 S_x^2 は x_i の分散、 $S_{x,y}$ は (x_i, y_i) の共分散です。

回帰直線

回帰直線とはモデルによる目的変数の推測値を直線で示したものです。

$$y = b_0 + b_1 \cdot x$$

データサイエンス発展I 教材例



DS発展II 各回の内容 (令和7年度予定)

1. 機械学習とニューラルネットワーク
2. ディープラーニング
3. 機械学習のアルゴリズム
4. 画像データ分析1: 画像処理
5. 画像データ分析2: 畳み込みニューラルネットワーク
6. 画像分類問題に取り組む1: データの確認とモデル化の準備
7. 画像分類問題に取り組む2: モデルの作成
8. テキストデータ分析1: 自然言語処理の基礎
9. テキストデータ分析2: ニューラルネットワーク
10. 課題解決に挑戦1: 課題設定と分析手法の決定
11. 課題解決に挑戦2: 分析に取り組む
12. 課題解決に挑戦3: モデル化に取り組む
13. 課題解決に挑戦4: コンペに挑戦
14. 課題解決に挑戦5: 発表準備
15. 課題解決に挑戦6: 最終発表会

データサイエンス発展II 授業の実施について

- 実習室にて対面授業を実施
- 教材: 第1回-第10回はeラーニング教材 (Signate) による実習, 最終課題として2つの課題から選択 (令和6年度)

- 最終課題:
 1. モノクロ顔画像の感情分類
 2. 青森県産りんごの価格予測

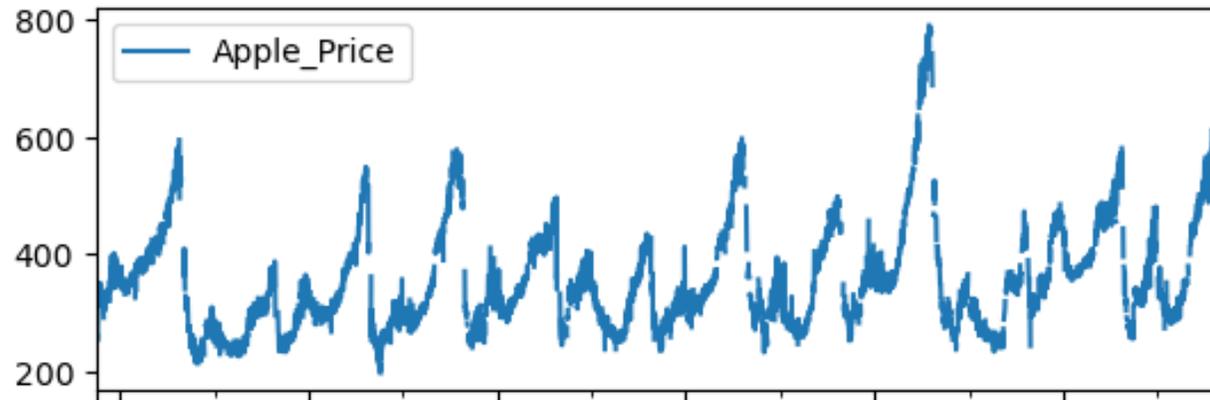
- 評価: eラーニング教材と最終課題の成果

最終課題：青森県産りんごの価格予測

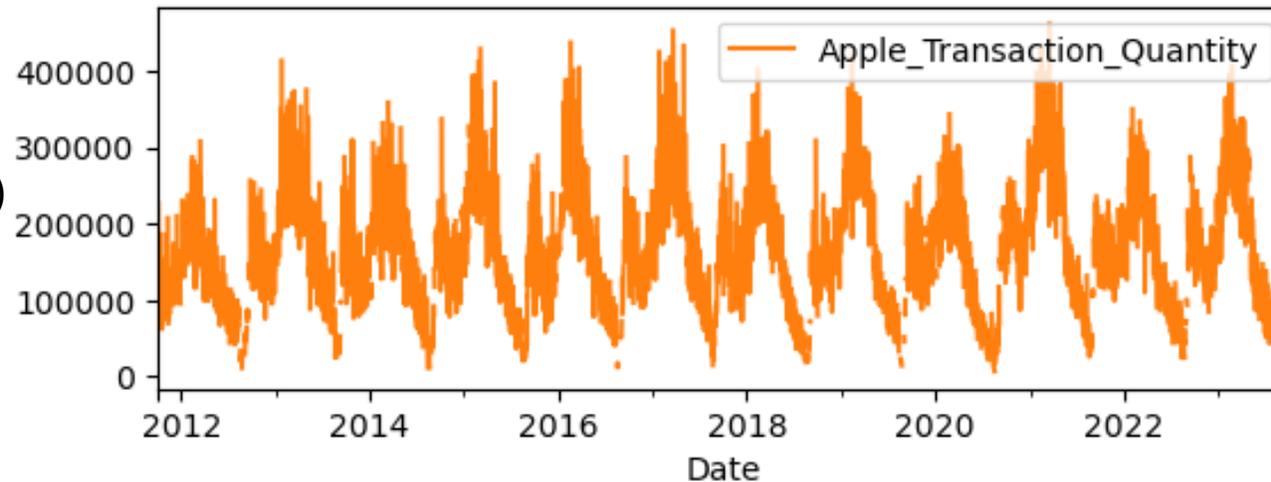
- 気温や日照，取引量などのデータを利用して**青森県産りんごの東京都中央卸売市場での卸売価格**を予測する時系列モデルや多変量回帰モデルを構築
- 消費地におけるりんご価格は，青森県のりんご産業に非常に重要
- 使用可能なデータ：東京都中央卸売市場におけるりんご，みかん，梨の卸売価格と取引量，弘前市の気象データ，東京の気象データ
- 予測：2023年10月1日から2024年9月30日（366日間）のりんご価格を予測

りんごの卸売価格・取引量推移

1 kgあたり
卸売価格 (円)



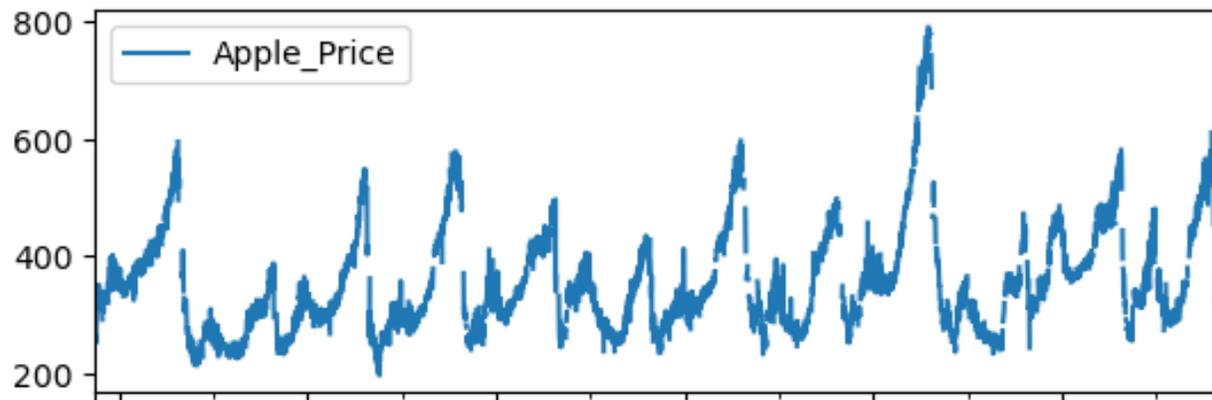
取引数量 (kg)



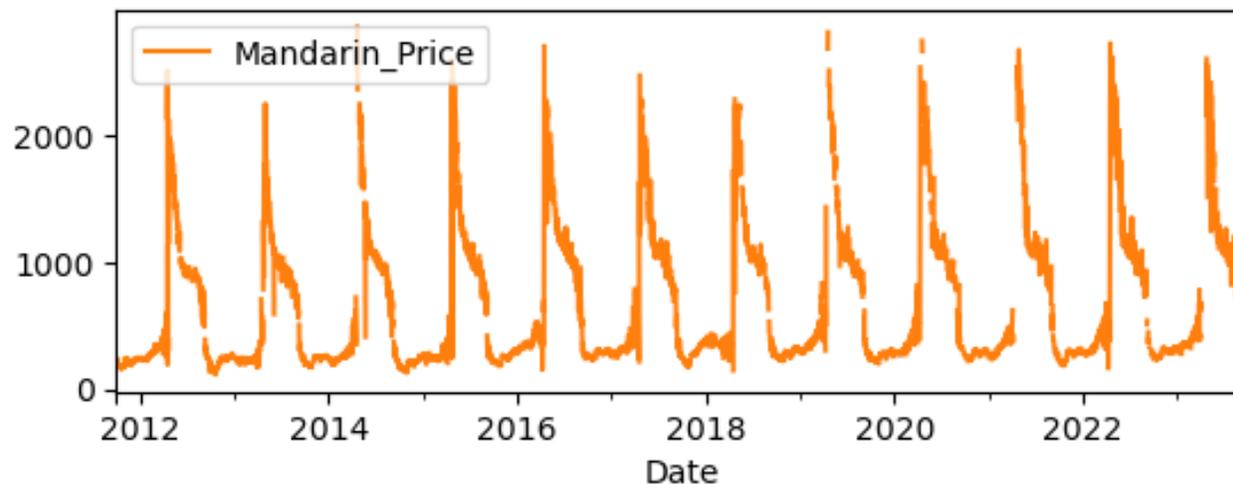
- 夏には卸売価格が上昇，取引量は減少する季節変動
- 近年卸売価格は上昇傾向，取引量は減少傾向

りんごとみかん価格の比較

りんご
1 kgあたり
卸売価格 (円)



みかん
1 kgあたり
卸売価格 (円)



- 他の果実 (例: みかん) と比較すると, 価格変動は不規則で, 予測が難しそうに見える

弘前大学での今後の展開

- 内製化教材をもちいた全学における授業実施
- 履修者の増加に対応するために、オンデマンド形態の授業を準備 (DS発展I)
 - オンデマンド授業用教材開発
 - 動画教材・小テストの作成, 演習教材の改良
- データサイエンス科目群の再編

まとめ

- データサイエンス発展I, データサイエンス発展IIを実施
 - **回帰分析**の手法と具体的な課題を用いて, データサイエンスならびに人工知能による実践的な課題解決の方法を学ぶ
 - **AIのアルゴリズム**を学び, 具体的な課題を用いてデータサイエンスならびに人工知能による実践的な課題解決の方法を学ぶ
- 内製化教材, オンデマンド教材を作成
- データサイエンス科目群の再編