

基礎演習

第5回 データ解析基礎（2）基本統計量

メディア情報コース
平居 悠（ひらい ゆたか）

到達目標

- データ解析の基本的なスキルを身につける。
- 適切に文献を引用し、自分の意見を記述したレポートを作成できる。
- 調べてわかったことをわかりやすく他者にわかりやすく伝えることができる。

前回

- 第1回 授業の概要、自己紹介 (4/13)
- 第2回 天文学入門 138億年の宇宙の歴史 (4/20)
- 第3回 図書館ツアー、データベース検索 (4/27)
- 第4回 **データ解析基礎 (1) プログラミング言語Python入門 (5/11)**
- 第5回 データ解析基礎 (2) 基本統計量 (5/18)
- 第6回 データ解析実習 (1) 図の作成 (5/25)
- 第7回 データ解析実習 (2) データの解釈 (6/1)
- 第8回 レポートの作成 (1) 文章の読み方と要約、レポートの構成と表現 (6/8)
- 第9回 レポートの作成 (2) 著作権 (6/15)
- 第10回 プレゼンテーションの方法 (6/22)
- 第11回 発表準備 (1) 発表資料の作成 (6/29)
- 第12回 発表準備 (2) 発表練習 (7/6)
- 第13回 発表会 (7/13)

前回の目標

プログラミング言語Pythonで
簡単なプログラムを作成でき
るようになる。

データ解析とは

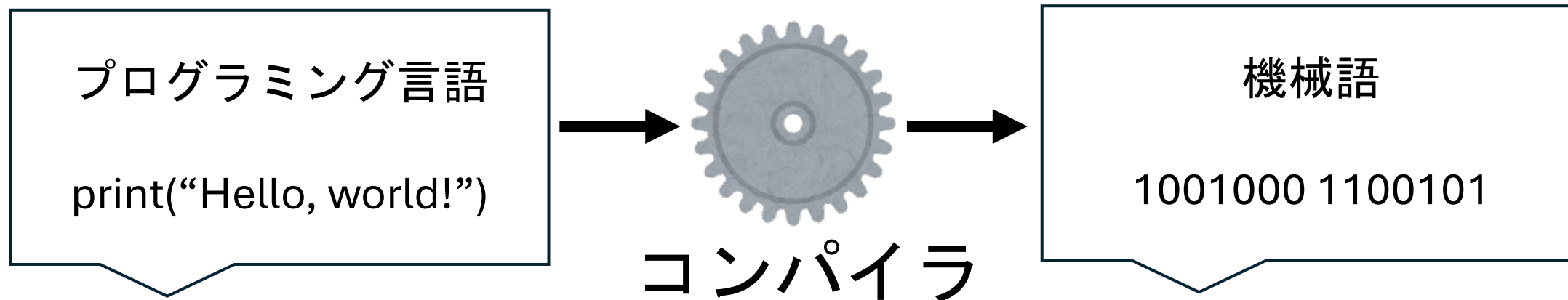
収集されたデータを分析した結果から、一定の法則性や共通点を見出し、問題の解決に役立てること

プログラミングとは？

プログラミング言語の
文法に従って処理手順
を書く工程

コンパイル

プログラミング言語をコンピュータが読める機械語に翻訳



主なプログラミング言語

Python
(パイソン)



開発者：
ガイド・ヴァンロッサム
(1991年)

用途：
データ解析、人工知能
など

授業：
基礎演習

Ruby
(ルビー)



まつもとゆきひろ
(1995年)

Web開発など

基礎プログラミング

C言語
(シーげんご)



デニス・リッチー
(1972年)

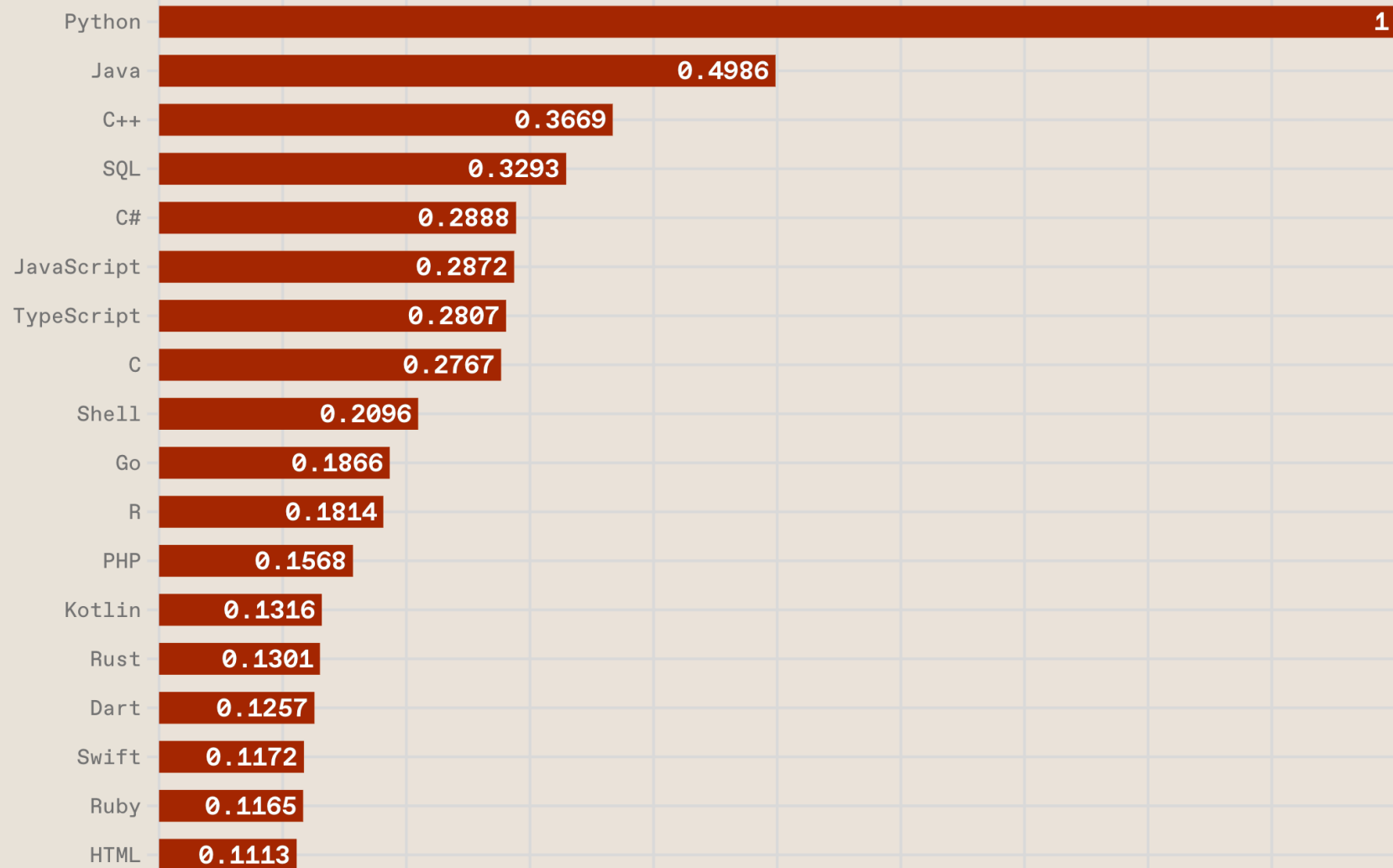
科学技術計算、OSなど

応用プログラミング、
データ構造とアルゴリズム

Top Programming Languages 2025

Click a button to see a differently weighted ranking

Spectrum Trending Jobs



なぜPythonを学ぶのか？

IEEEプログラミング言語ランキング9年連続1位
(2017-2025年)

なぜPythonを学ぶのか？

データ解析や人工知能が得意

充実したライブラリ（再利用可能なコードの集合体）

データ分析：Numpy, Pandas, Matplotlibなど

機械学習：Scikit-learn, TensorFlowなど

Web開発：Flask, Djangoなど

なぜPythonを学ぶのか？

文法（書き方）がシンプル

例：“Hello, world!”と出力するプログラム



```
print("Hello, world!")
```



```
#include <stdio.h>

int main (void)
{
    printf("Hello, world!\n");
    return 0;
}
```

Pythonインストール

Pythonのダウンロード

<https://www.python.org/downloads/>



The screenshot shows the Python.org website interface. At the top, there is a navigation bar with links for Python, PSF, Docs, PyPI, Jobs, and Community. Below this is the Python logo and a search bar with a 'GO' button and a 'Socialize' link. A secondary navigation bar contains links for About, Downloads, Documentation, Community, Success Stories, News, and Events. The main content area features a prominent heading 'Download the latest version for macOS' and a yellow button labeled 'Download Python 3.14.4'. Below the button, there are links for other operating systems: 'Looking for Python with a different OS? Python for [Windows](#), [Linux/Unix](#), [macOS](#), [Android](#), [other](#)'. Further down, there are links for 'Want to help test development versions of Python 3.15? [Pre-releases](#), [Docker images](#)'. The background of the page features an illustration of two parachutes with cargo boxes hanging from them, set against a blue sky with clouds.

Numpy

数値計算に特化したライブラリ

大規模データを効率良く扱える

公式サイト：

<https://numpy.org/ja/>



Pandas

データ整形用ライブラリ

機能：

- ファイルの読み書き
- 列・行の追加や削除
- データの抽出 など



公式サイト：

<https://pandas.pydata.org/>

Matplotlib



グラフ描画用ライブラリ

データの可視化に便利

公式サイト：

<https://matplotlib.org/>

JupyterLab

対話型のプログラム実行環境

Webブラウザ上で各種プログラムの実行と結果の参照、ドキュメント作成などが行える

他の言語 (Julia, R, Rubyなど)にも対応

公式サイト：

<https://jupyter.org/>



JupyterLabの実行

PowerShellに

“`py -m jupyter lab`”

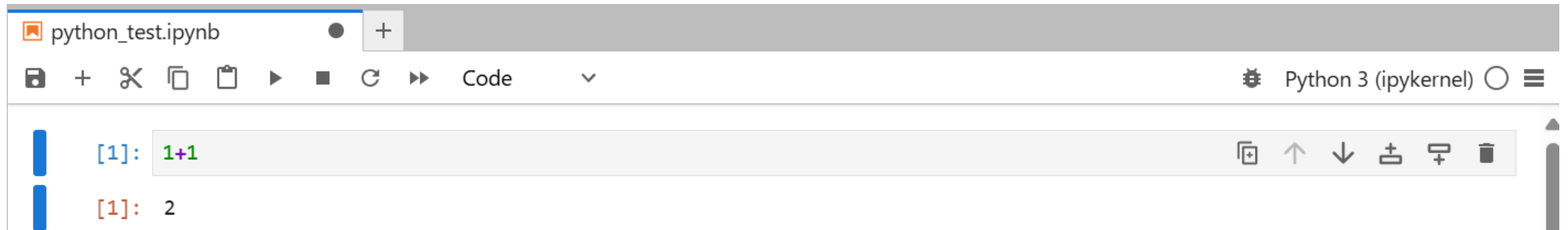
と入力してEnterキーを

押す

Pythonを実行してみよう

1. セルに“1+1”と入力

2. Shift+Enterか▶▶をクリック



The screenshot shows a Jupyter Notebook window titled 'python_test.ipynb'. The interface includes a toolbar with icons for saving, adding, deleting, copying, pasting, and running code. The code cell contains the text `[1]: 1+1`. Below the code cell, the output is displayed as `[1]: 2`. The notebook is running on Python 3 (ipykernel).

今回

- 第1回 授業の概要、自己紹介 (4/13)
- 第2回 天文学入門 138億年の宇宙の歴史 (4/20)
- 第3回 図書館ツアー、データベース検索 (4/27)
- 第4回 データ解析基礎 (1) プログラミング言語Python入門 (5/11)
- 第5回 データ解析基礎 (2) 基本統計量 (5/18)**
- 第6回 データ解析実習 (1) 図の作成 (5/25)
- 第7回 データ解析実習 (2) データの解釈 (6/1)
- 第8回 レポートの作成 (1) 文章の読み方と要約、レポートの構成と表現 (6/8)
- 第9回 レポートの作成 (2) 著作権 (6/15)
- 第10回 プレゼンテーションの方法 (6/22)
- 第11回 発表準備 (1) 発表資料の作成 (6/29)
- 第12回 発表準備 (2) 発表練習 (7/6)
- 第13回 発表会 (7/13)

今回の目標

プログラミング言語Python
でデータを読み込んで基本
統計量を算出できるように
なる。

今回の内容

解析するデータについて
データの読み込み
基本統計量

今回の内容

解析するデータについて

データの読み込み

基本統計量

解析するデータ

天の川銀河形成シミュレーション

論文：

Hirai et al. (2022), MNRAS, 517, 4856

<https://academic.oup.com/mnras/article/517/4/4856/6808594>

Hirai et al. (2024), ApJ, 970, 105

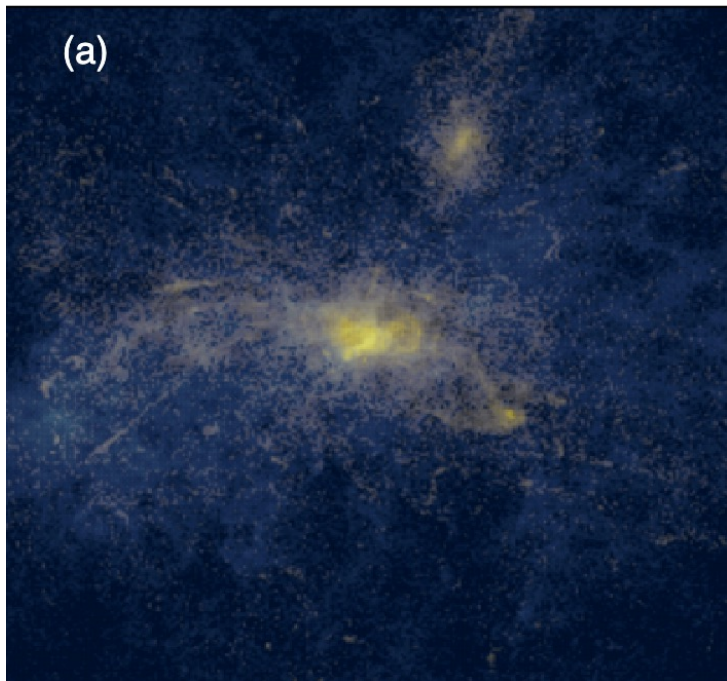
<https://iopscience.iop.org/article/10.3847/1538-4357/ad500c>

Hirai et al. (2025), ApJ, 9990, 125

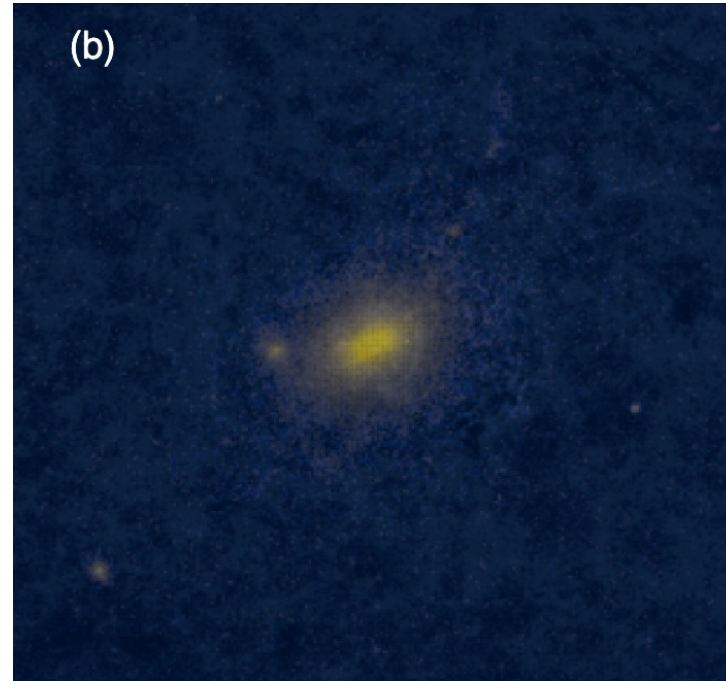
<https://iopscience.iop.org/article/10.3847/1538-4357/adf10a>

銀河形成シミュレーションとは？

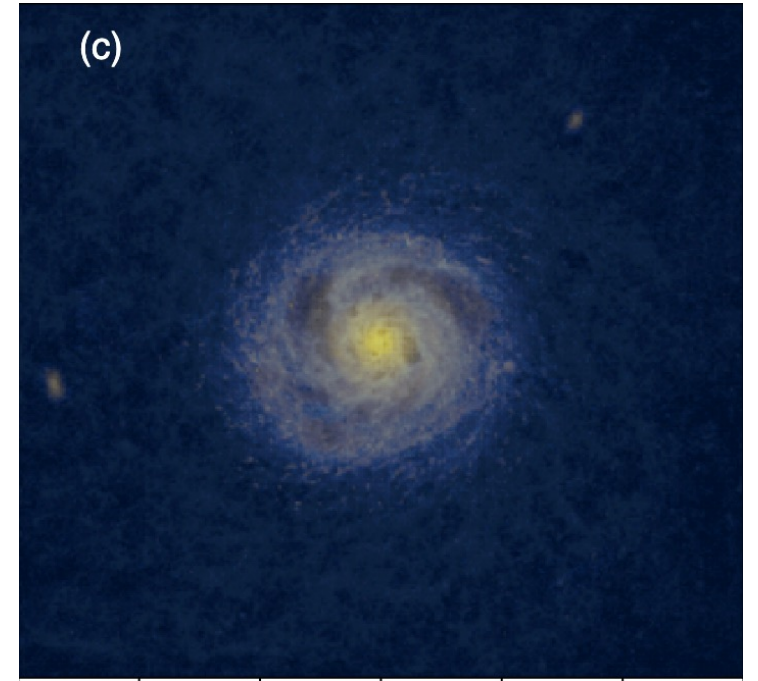
宇宙初期から現在まで銀河ができる過程を計算



宇宙誕生から
10億年後



50億年後



Hirai et al. (2025)

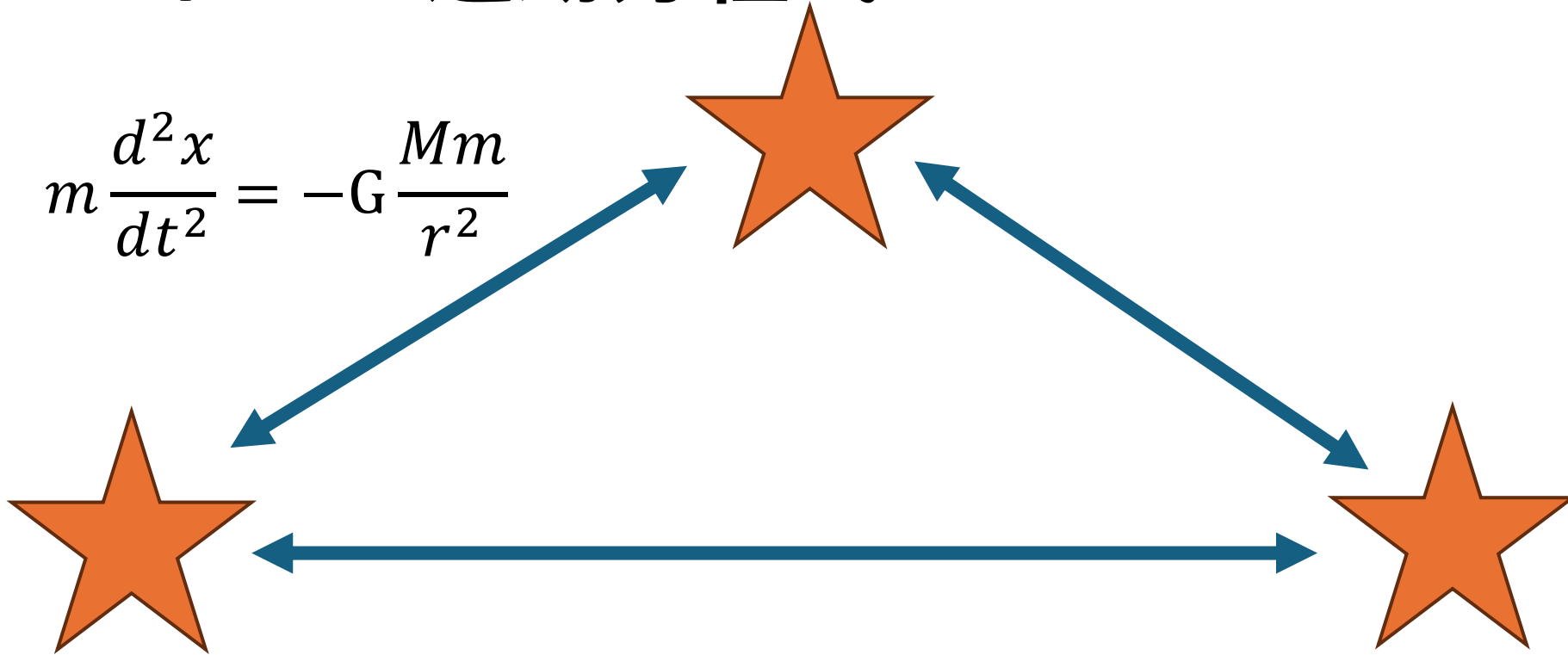
138億年後

何を計算しているのか？

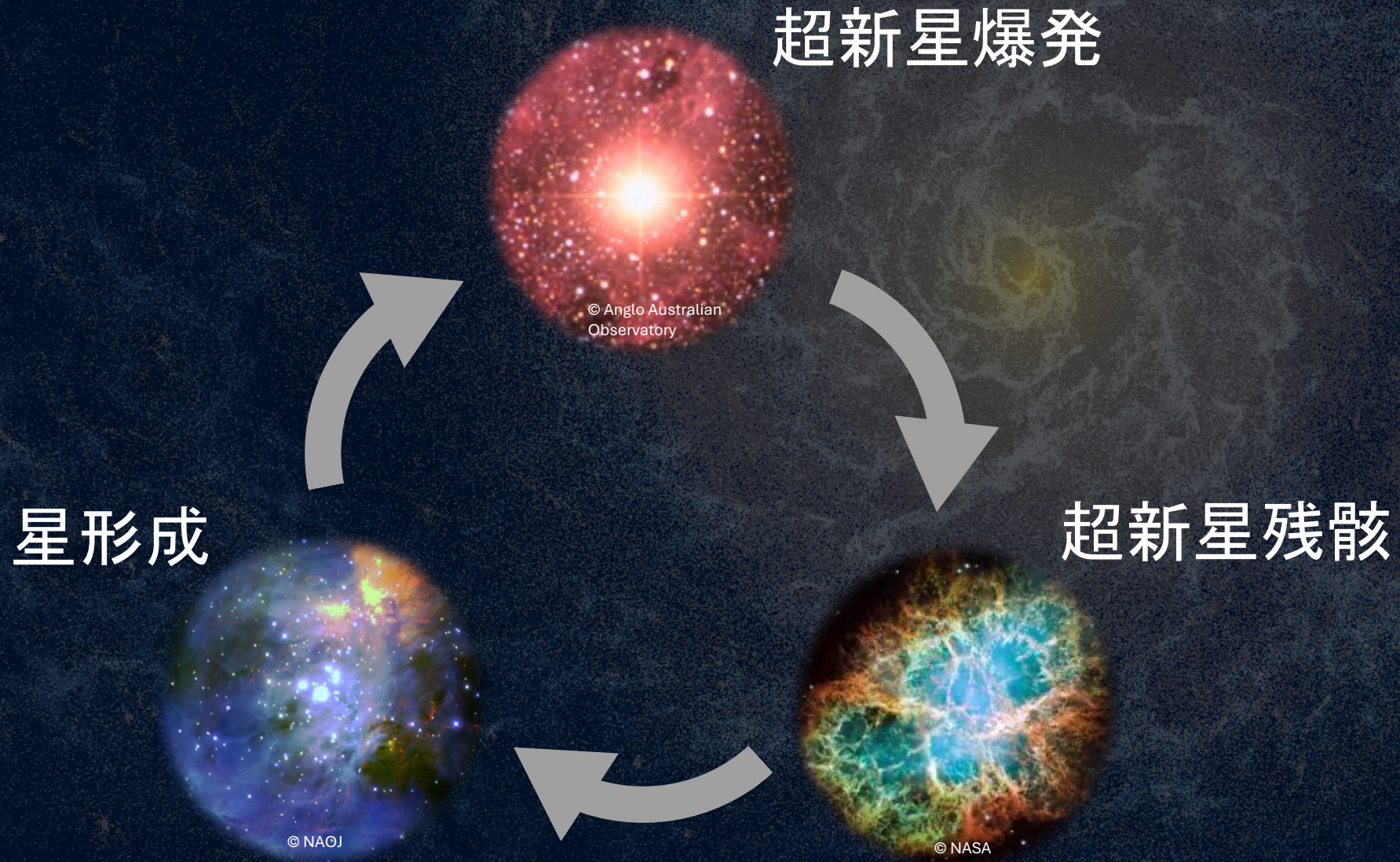
星やダークマターの重力相互作用

ニュートンの運動方程式

$$m \frac{d^2 x}{dt^2} = -G \frac{Mm}{r^2}$$



銀河内での物質循環（銀河の化学進化）



スーパーコンピュータ「アテルイIII」



基本的物理量

位置

単位はキロパーセク (kpc)

1 pc = 3.26 光年

1 kpc = 1000 pc

基本的物理量

星の生まれた時刻

単位は10億年 (Gyr)

基本的物理量

元素量

太陽と比べてどれくらいの元素量か？

鉄の場合

[Fe/H]

$= \log_{10}(\text{鉄の数密度} / \text{水素の数密度})$

$-\log_{10}(\text{太陽の鉄の数密度} / \text{太陽の水素の数密度})$

[Fe/H] = 0なら太陽と同じ、[Fe/H] = -1なら
太陽の1/10の鉄の量

シミュレーションデータからわかること

銀河の構造

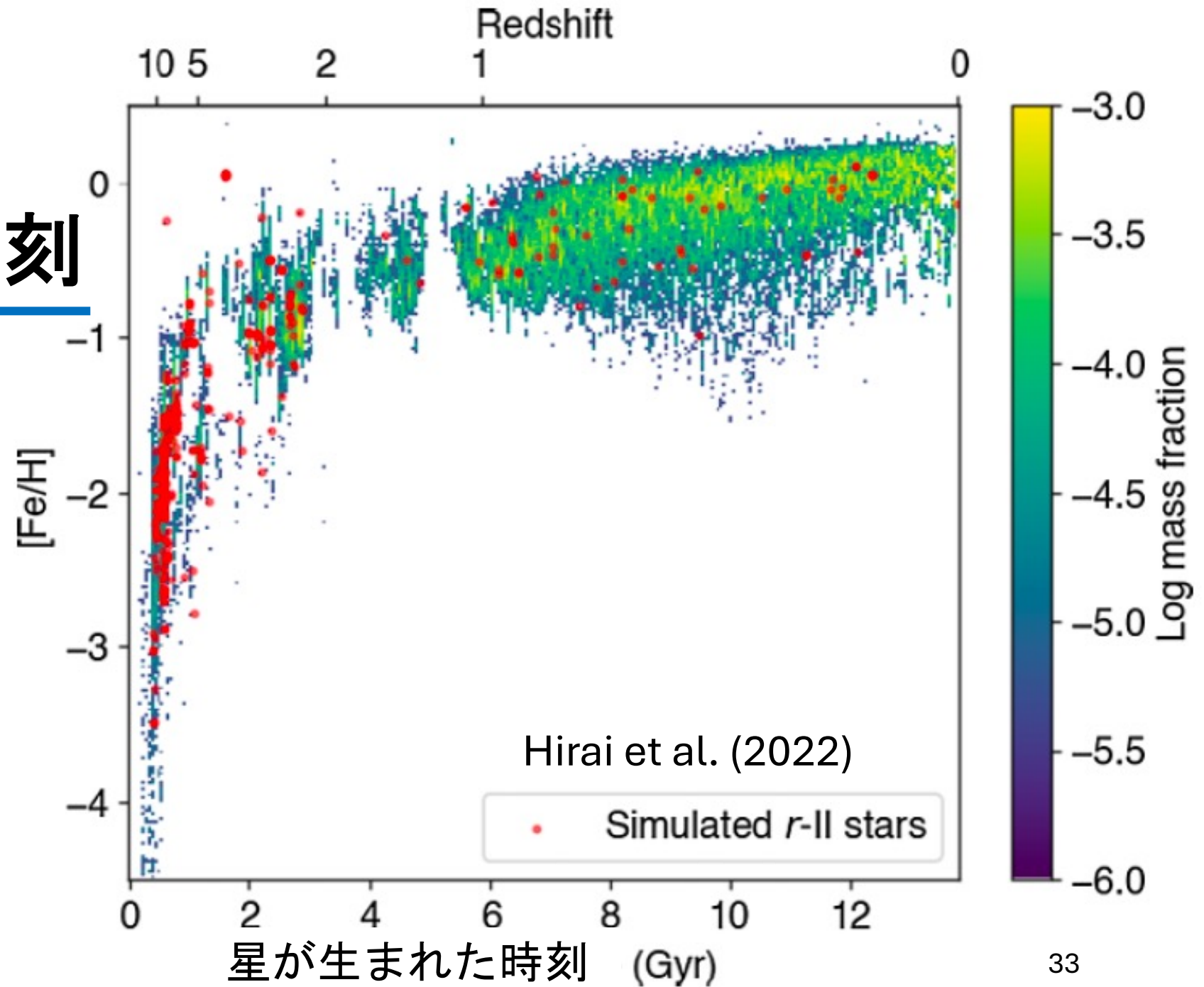


バルジ

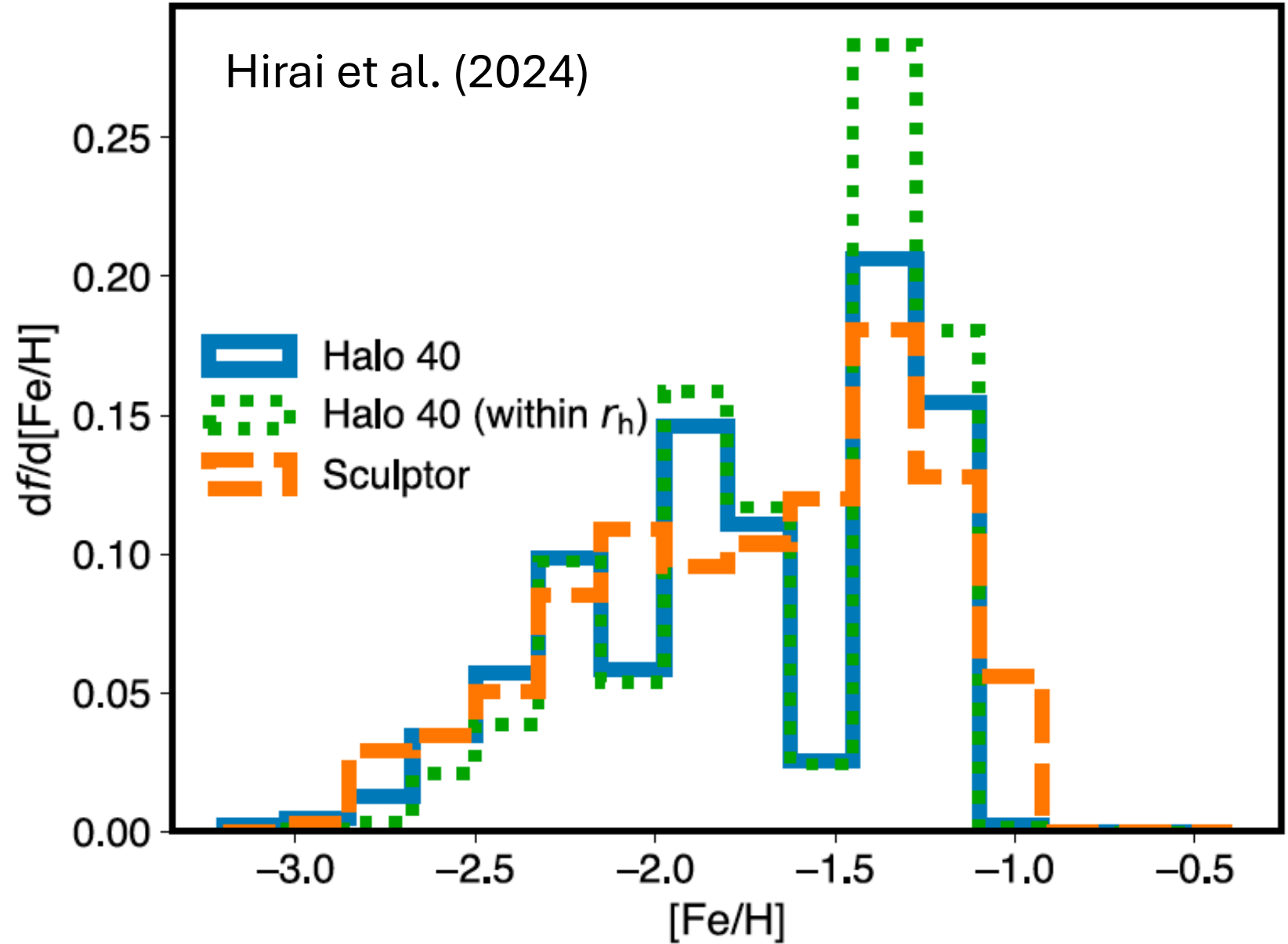
円盤

ハロー

星が生まれた時刻



金属量分布 (鉄の量の分布)

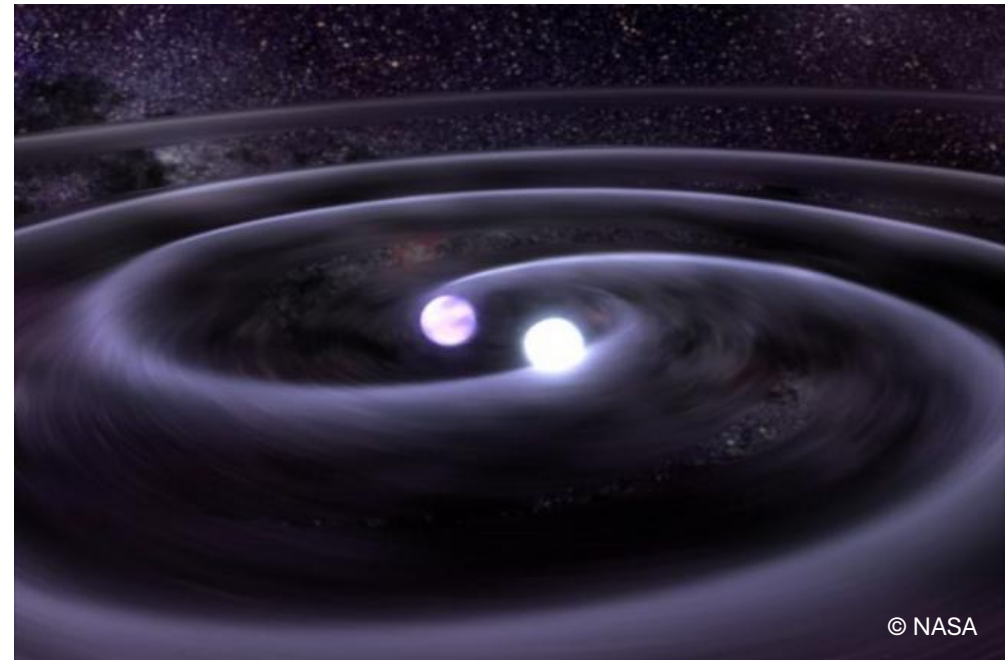


元素の起源、化学進化の速さ

酸素、マグネシウム、ケイ素など
重い星の超新星爆発で放出（約1000万年）

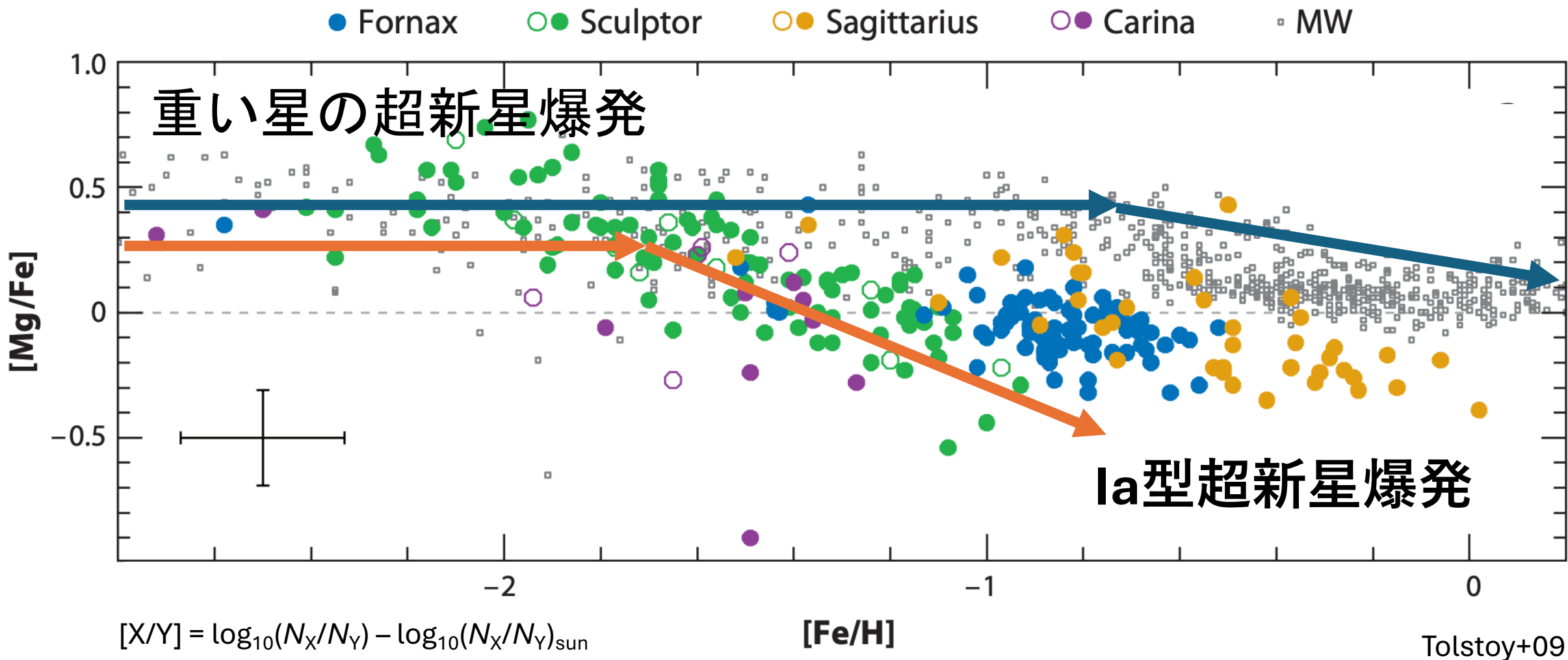
鉄

重い星の超新星爆発
+ 軽い星の連星の超新星爆発
（Ia型超新星爆発、1億年以上）



© NASA

元素の起源、化学進化の速さ



今回の内容

解析するデータについて

データの読み込み

基本統計量

データのダウンロード

http://roy.e.koeki-u.ac.jp/~hirai/2025/basic_skills/05/milkyway_data.csv

からダウンロード

レコードとフィールド

レコード：全ての項目が揃った1単位のデータ

フィールド：レコードの中にある項目

レコード

ID3	x	y	z	vx	vy	vz	t_form	MgH	FeH	MgFe
96282195	0.0908529	0.647045	1.096310000	-121.438	16.5403	66.8084	3.618454351	-0.6094579	-0.5087503	-0.1007076

フィールド

CSVファイルについて

複数のフィールドごとにカンマ
(,)で区切り、1レコードを1行に
おさめたファイル形式

ファイル内のデータの説明

ID3: 星のID

x, y, z: 位置 (kpc)

v_x, v_y, v_z: 速度 (km/s)

t_{form}: 星の形成時刻 (Gyr)

MgH: [Mg/H]

FeH: [Fe/H] (金属量)

MgFe: [Mg/Fe]

ダウンロードしたファイルの移動

1. PowerShellを開く
2. ディレクトリ“basic_skills”に移動

```
cd basic_skills
```

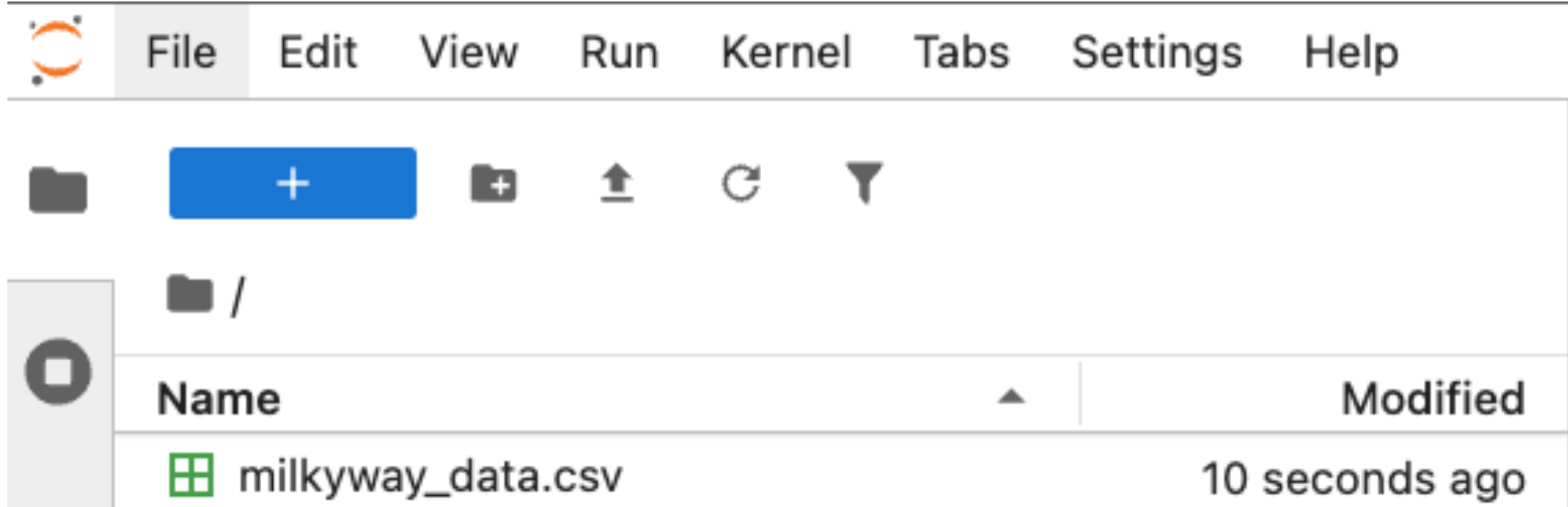
3. ファイルをダウンロードディレクトリから持ってくる

```
mv ../Downloads/milkyway_data.csv ./
```

4. “ls”と入力し、データを移動できたことを確認する。

新たなノートブックを生成

1. JupyterLabを開き、左上の+ボタンから新たなノートブックを作成。
2. ファイル名は“milkyway.ipynb”とする。



必要なライブラリのインポート

Pandas, Numpy, Matplotlibをインポートする。

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

データの読み込み

read_csv()を用いる

```
df = pd.read_csv("milkyway_data.csv")
```

データフレーム (DataFrame)

二次元の表形式のデータ

今回の内容

解析するデータについて
データの読み込み
基本統計量

最大値: maxメソッド

データの中で最大の値

一番若い星

一番高い金属量

```
df["t_form"].max()
```

```
df["FeH"].max()
```

```
13.800810339999998
```

```
0.7089017673474292
```

最小値: minメソッド

データの中で最小の値

一番古い星

一番低い金属量

```
df["t_form"].min()
```

```
df["FeH"].min()
```

```
0.33695625943
```

```
-3.689660981820394
```

最頻値: mode メソッド

データの中で最も出現回数が多い値

```
df["t_form"].mode()
```

0	0.465165
1	0.781093
2	1.323300
3	2.083576
4	10.528863
5	12.325295
6	13.576110

これらの時刻にできた星が最も多い

平均値: mean メソッド

個々の数値を全て足し合わせた後、
数値の個数で割った値

星の形成時刻の平均値 星の金属量の平均値

```
df["t_form"].mean()
```

```
6.673826215048475
```

```
df["FeH"].mean()
```

```
-0.35654102618473477
```

データのばらつき

偏差：各データ値と平均値の差

分散：偏差の2乗を足し合わせて、データの個数で割った値

星の形成時刻の分散

```
df["t_form"].var()
```

14.96144099389696

星の金属量の分散

```
df["FeH"].var()
```

0.260732480749871

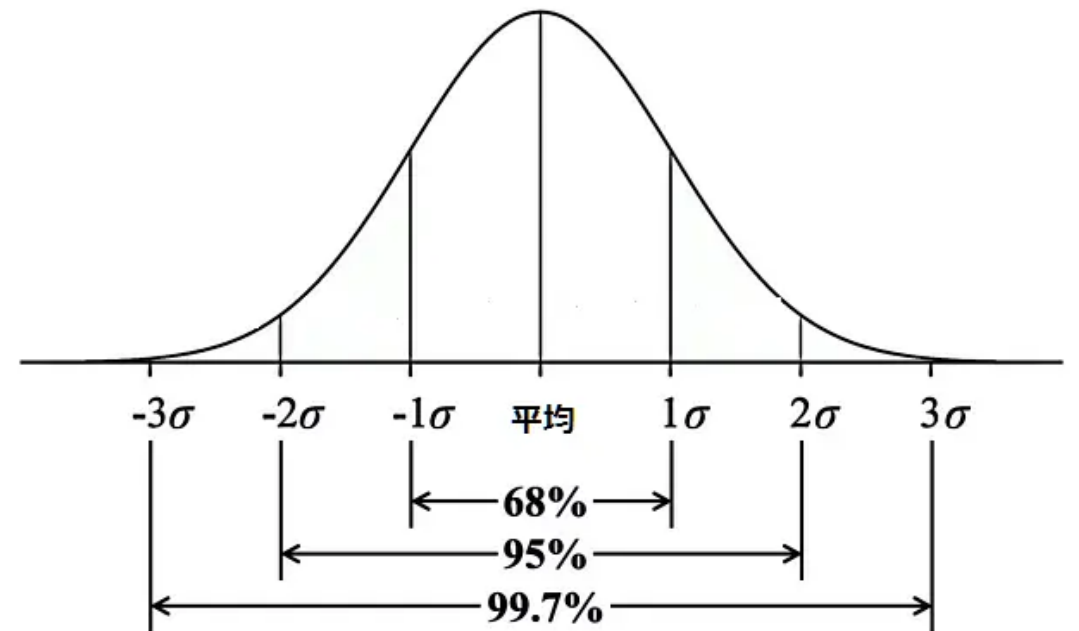
データのばらつき

標準偏差：

分散の平方根をとった値

標準偏差の意味：

平均値50、標準偏差10の場合、平均値50に対して、各データが標準的に10離れている。50 ± 10の範囲に全データの約68%が含まれる。



<https://avilen.co.jp/personal/knowledge-article/standard-deviation/>

データのばらつき：標準偏差

形成時刻の標準偏差

```
df["t_form"].std()
```

```
3.8680021967285594
```

金属量の標準偏差

```
df["FeH"].std()
```

```
0.5106197026651743
```

データのばらつき

偏差値：分布の平均が50、標準偏差が10となるように変換した数値

$$\text{偏差値} = ((\text{個々のデータ値} - \text{平均値}) / \text{標準偏差}) * 10 + 50$$

偏差値の計算

シミュレーションデータ内で[Fe/H]
=-1の星の偏差値を求めよう

```
mean = df["FeH"].mean()  
std = df["FeH"].std()  
feh = -1
```

```
deviation =  
deviation_value = #偏差値
```

中央値: median メソッド

データを小さい順に並べたときにちょうど真ん中に来る値

星の形成時刻の中央値

```
df["t_form"].median()
```

6.47567872935

星の金属量の中央値

```
df["FeH"].median()
```

-0.2815630059364782

四分位数: quantile() メソッド

データを小さい順に並べたとき、データの数で4等分した時の区切り値

第一四分位数：小さい方から25%

第二四分位数：中央値

第三四分位数：小さい方から75%

四分位範囲 = 第三四分位数 - 第一四分位数

四分位数: quantile() メソッド

金属量の第一四分位数

```
df["FeH"].quantile(0.25)
```

```
-0.613387421732575
```

第三四分位数

```
df["FeH"].quantile(0.75)
```

```
-0.011540741636342174
```

四分位範囲

```
df["FeH"].quantile(0.75) - df["FeH"].quantile(0.25)
```

今回の目標

プログラミング言語Python
でデータを読み込んで基本
統計量を算出できるように
なる。

今回の内容

解析するデータについて
データの読み込み
基本統計量

次回

- 第1回 授業の概要、自己紹介 (4/13)
- 第2回 天文学入門 138億年の宇宙の歴史 (4/20)
- 第3回 図書館ツアー、データベース検索 (4/27)
- 第4回 データ解析基礎 (1) プログラミング言語Python入門 (5/11)
- 第5回 データ解析基礎 (2) 基本統計量 (5/18)
- 第6回 **データ解析実習 (1) 図の作成 (5/25)**
- 第7回 データ解析実習 (2) データの解釈 (6/1)
- 第8回 レポートの作成 (1) 文章の読み方と要約、レポートの構成と表現 (6/8)
- 第9回 レポートの作成 (2) 著作権 (6/15)
- 第10回 プレゼンテーションの方法 (6/22)
- 第11回 発表準備 (1) 発表資料の作成 (6/29)
- 第12回 発表準備 (2) 発表練習 (7/6)
- 第13回 発表会 (7/13)