

基本操作

df.列1	列1を抽出
df.[“列1”]	列1を抽出
df.[[“列1”, “列2”, “列3”]]	列1~3を抽出
df.loc[行ラベル, 列ラベル]	指定した行列ラベルを取得
df.iloc[行番号, 列番号]	指定した行列番号を取得

概要確認

df.head() or df.tail()	先頭or末尾から5行取得
df.info()	データの概要を取得
df.shape	行数、列数を取得
df.max() df.min()	列の最大値、最小値を取得
df.mean() df.std()	列の平均値、標準偏差を取得
df.index	インデックスを取得
df.columns	列名を表示

ソート、置換、結合

df.sort_values(“列1”)	列1を昇順で並べ替え
df.sort_values(“列1”, ascending=False)	列1を降順で並べ替え
df.replace({1:100})	1を100に置換
pd.concat([df1, df2])	df1とdf2を結合

欠損値処理

df.dropna()	欠損値がある行を削除
df.isnull()	欠損値がある行を抽出
df.fillna(100)	欠損値を100で穴埋め

ファイルの読み込み、書き出し

pd.read_csv(“ファイル名”)	ファイルの読み込み
df.to_csv(“書き出しファイル名”)	ファイルの書き出し

※引数で文字コードやヘッダーの指定が可能

その他のメソッド

<code>pd.get_dummies(df[["列1","列2"]])</code>	列1、列2のカテゴリでダミー変数化
<code>pd.to_numeric("列1")</code>	文字型の要素を持つ列1を数値型に変換
<code>df.sample()</code>	ランダムに1行を取得