

問2. 以下の問いに答えよ。なお、解答の根拠となるipynb/htmlファイルが両方とも提出されていない場合は、解答を無効とする。(30点)

cs3-mid-2-1.csvとcs3-mid-2-2.csvを読み込んで一連の処理を行うノートブックの (1) などの空欄を埋め、また問いに答えよ。空欄を埋めるプログラムは1行で答えること。なお、提出するipynb/htmlファイルには、以下に表示されていないプログラム行が含まれていても構わない。

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

cs3-mid-2-1.csvの全データをデータフレームdf1、cs3-mid-2-2.csvの全データをデータフレームdf2に読み込み(解答不要)。

df1の重複行をすべて表示(1)。  
重複を削除して、行番号を0からの連番に振り直し、変数df1mに代入(2)。(元の行番号を格納する列は生成させない)  
df1mの行数(データ数)を答えよ(3)。

```
display( (1) )
(2)
# num: (3)
```

(1)

(2)

(3)

df1mの各列の欠損値の数を表示(4)。欠損値を1つ以上含む列をすべて答えよ(5)。次の例のように、‘や’は付けず、複数ある場合はカンマで区切ること(例 a1, a2, a3)。  
df1mの欠損値を1つでも含む行を表示(6)。  
df1mの欠損値を1つでも含む行を削除して、行番号を0からの連番に振り直し、変数df1m2に代入(7)。

```
print( (4) )
# columns with missing values: (5)
display( (6) )
(7)
```

(4)

(5)

(6)

(7)

df2のc2列に含まれる値とその個数を表示せよ(8)。Xの個数を答えよ(9)。c2列の値yをすべてYに置換する(10)。

```
display( (8) )
# number of X: (9)
(10)
```

(8)

(9)

(10)

df1m2の各行の右側に、df2の対応する行を結合し、結果を変数df3に代入(11)。このとき、df1m2のID列の値が、df2のidx列の値と一致するようにし、データがID列とidx列の両方に共通して現れる場合のみがdf3に現れるようにする。

```
(11)
```

(11)

df3のデータフレームをCSVファイルmid-p2-out.csvに保存(12)。なお、行番号をindex列に保存する必要はない。またencoding=など、その他のオプションの設定は不要。また、df3の行数(データ数)と列数を(13)および(14)に答えよ。

```
(12)
# num: (13), (14)
```

(12)

(13)

(14)

df3のc1列とc2列を使ってクロス集計表を作成する(15)。c1の値がB、c2の値がXのデータ数を答えよ(16)。

```
display( (15) )
# number of data: (16)
```

(15)

(16)