

1 複数のデータセットの結合

【ファイル】

→ 【2つのデータセットを結合する】

によって**複数のデータセットを結合**した新たなデータセットを作成することができる。しばしば必要になるのは、同じ構造のデータセットを組み合わせる全体の大きなデータセットを作成するような状況（例えば施設Aと施設Bの患者データを統合する）であるが、その場合は「**行単位での結合**」を選択すればよい。以下の例ではData2にだけWBCという項目が含まれているが、「共通の行または列のみ結合」のオプションを選ぶと共通して含まれる項目だけが新しく統合されるデータセットに含まれるようになり、このオプションを選ばなければこの項目も欠損値はNAとして新しいデータセットに含まれる。

Data1

	Age	Sex	Disease	LDH	ID
1	24	M	AML	542	12345-1
2	34	F	AML	184	23456-1
3	31	M	ALL	945	34567-1
4	56	M	AML	235	45678-1
5	28	F	ALL	412	56789-1

Data2

	Age	Sex	Disease	LDH	WBC
1	43	F	ALL	345	32400
2	64	M	AML	234	9800
3	38	M	AML	312	2600
4	16	F	AML	216	14500

2つのデータセットを結合する

結合したデータセット名: MergedDataset

1番目のデータセット (1つ選択) 2番目のデータセット (1つ選択)

Data1 Data2

結合の方法

☒ 行単位での結合 ☐ 共通の行または列のみ結合

列単位での結合の照合方法

☒ 行番号で照合する ☐ 指定した変数で照合する

ヘルプ OK キャンセル

MergedDataset

	Age	Sex	Disease	LDH	ID	WBC
1	24	M	AML	542	12345-1	NA
2	34	F	AML	184	23456-1	NA
3	31	M	ALL	945	34567-1	NA
4	56	M	AML	235	45678-1	NA
5	28	F	ALL	412	56789-1	NA
11	43	F	ALL	345	<NA>	32400
21	64	M	AML	234	<NA>	9800
31	38	M	AML	312	<NA>	2600
41	16	F	AML	216	<NA>	14500

同様に、同じサンプルに対して異なる項目のデータを含むデータセットを組

み合わせて新しいデータセットを作成する場合は「**列単位での結合**」を選択する。「共通の行または列のみ結合」のオプションを選ぶと共通して含まれるサンプルだけが新しく統合されるデータセットに含まれるようになるが、このオプションを選ばなければすべてのサンプルが欠損値は NA として新しいデータセットに含まれる。行の照合方法は「行番号で照合する」方法と「指定した変数で照合する」方法が選択できる。前者の場合は **row.names** が同じサンプルが同一サンプルとみなされるため、特に **row.names** を指定していない場合は上からの行数が同じサンプルが同一サンプルとして統合される。一方、後者の場合は、たとえば ID のように各サンプルに固有のデータが存在するなら、その変数を利用して確実にサンプルを照合して列を結合することができる。照合する変数に重複があると総当たりの組み合わせで行が作成される。なお、両者のスクリプト上の違いは前者が **merge (データセット 1, データセット 2, all=TRUE, by="row.names")** で、後者は **by="row.names"** の部分が **by.x="ID", by.y="ID"** となる。この **all=TRUE** の部分を **all.x=TRUE** とすれば

	Age	Sex	Disease	LDH	ID
1	24	M	AML	542	12345-1
2	34	F	AML	184	23456-1
3	31	M	ALL	945	34567-1
4	56	M	AML	235	45678-1
5	28	F	ALL	412	56789-1

	WBC	Hb	ID	Disease
1	3200	12.6	23456-1	AML
2	12480	7.5	56789-1	ALL
3	6450	10.4	45678-1	AML
4	21250	9.8	67890-1	ALL
5	1480	11.2	12345-1	AML

列を結合するための変数

列を結合するために照合する変数(1番目のデータセット) 列を結合するために照合する変数(2番目のデータセット)

Age	Disease
Disease	Hb
ID	ID
LDH	WBC
Sex	

OK キャンセル

	ID	Age	Sex	Disease.x	LDH	WBC	Hb	Disease.y
1	12345-1	24	M	AML	542	1480	11.2	AML
2	23456-1	34	F	AML	184	3200	12.6	AML
3	34567-1	31	M	ALL	945	NA	NA	<NA>
4	45678-1	56	M	AML	235	6450	10.4	AML
5	56789-1	28	F	ALL	412	12480	7.5	ALL
6	67890-1	NA	<NA>	<NA>	NA	21250	9.8	ALL