

ใบงานที่ 4

1. จากข้อมูลตารางที่ 1 ตารางที่ 2 และคำอธิบายข้อมูลในตารางที่ 3 ให้นักศึกษาอธิบายการทำความสะอาดข้อมูล และการผสมผสานข้อมูลตารางที่ 1 และตารางที่ 2 เข้าด้วยกัน พร้อมเขียนคำสั่งในโปรแกรม R

ตารางที่ 1 ข้อมูลบุคคล

perid_pre	titme	age	gender	hholdid	comment
7533	นาย	25	1	94020180318	3สมาชิกในครัวเรือน
5659	นางสาว	27		94090184819	3สมาชิกในครัวเรือน
1497	1	4		94030212247	2เด็กอายุ 3-25 ปี
2174	2		2	11	2เด็กอายุ 3-25 ปี
8408	นางสาว	38	2	94100181193	3สมาชิกในครัวเรือน
10185	นาย	37	1	94010322560	3สมาชิกในครัวเรือน
9435	เด็กชาย	3	1	94120040453	3สมาชิกในครัวเรือน
2691	3	16	1	94030176003	2เด็กอายุ 3-25 ปี
2575	1	12		94020240159	2เด็กอายุ 3-25 ปี
3105	1	11	1	94120045714	2เด็กอายุ 3-25 ปี
7416	นางสาว	33		94020127654	3สมาชิกในครัวเรือน
2271	3	15		94010275308	2เด็กอายุ 3-25 ปี
2980	2	15	2	94100278250	2เด็กอายุ 3-25 ปี
8091	เด็กหญิง	5	2	94090177804	3สมาชิกในครัวเรือน
9762	เด็กชาย	12	1	94100224729	3สมาชิกในครัวเรือน

ตารางที่ 2 ข้อมูลการศึกษาและการทำงาน

perid_pre	a9	a10	a11	a12	a13	a14
5659	ได้	ได้	ม.ปลายหรือเทียบเท่า	ไม่ได้เรียน		ทำงาน
7416	ได้	ได้	ม.ปลายหรือเทียบเท่า	ม.ปลายหรือเทียบเท่า	ไม่ได้เรียน	ทำงาน
7533	ได้	ได้	ม.ปลายหรือเทียบเท่า	ม.ปลายหรือเทียบเท่า	ไม่ได้เรียน	ทำงาน
8091	ได้	ได้		ประถมศึกษา	ไปเรียนสม่ำเสมอ	ไม่ทำงาน
8408	ได้	ได้	ม.ปลายหรือเทียบเท่า			ทำงาน
9435	ได้	ได้	ต่ำกว่าประถม	ต่ำกว่าประถม	ไปเรียนสม่ำเสมอ	ไม่ทำงาน
9762	ได้	ได้	ประถมศึกษา	ประถมศึกษา	ไปเรียนสม่ำเสมอ	
10033	ได้	ได้	ปวส.หรืออนุปริญญา	ไม่ได้เรียน		ทำงาน
10185	ได้	ได้	ม.ต้นหรือเทียบเท่า	ไม่ได้เรียน	ไม่ได้เรียน	ทำงาน
10296	ได้	ได้	ม.ปลายหรือเทียบเท่า			ทำงาน
9745	ได้	ได้	ไม่ได้เรียน	ไม่ได้เรียน	ไม่ได้เรียน	
5232	ได้	ได้	ม.ปลายหรือเทียบเท่า	ไม่ได้เรียน	ไม่ได้เรียน	ทำงาน
9752	ได้	ได้	ประถมศึกษา	ไม่ได้เรียน	ไม่ได้เรียน	ไม่ทำงาน
6038	ได้	ได้	ม.ปลายหรือเทียบเท่า	ไม่ได้เรียน	ไม่ได้เรียน	

ตารางที่ 3 คำอธิบายข้อมูลการศึกษาและการทำงาน

a9	การอ่านภาษาไทย(0=ไม่ได้,1=ได้)
a10	การเขียนภาษาไทย(0=ไม่ได้,1=ได้)
a11	การศึกษาสูงสุด(เลือก0-8)
a12	กำลังศึกษาระดับ(เลือก0-9)
a13	สถานภาพการศึกษา(เลือก0-3)
a14	สถานภาพการทำงาน(เลือก0-1)

คำอธิบายการทำความสะอาดข้อมูล และการผสานข้อมูล

ตารางที่1 ข้อมูลบุคคล ทำความสะอาด และผสานข้อมูลจากตารางที่ต่างกันเข้าด้วยกัน ของตัวแปร titnme โดยที่ เด็กชาย = 1, เด็กหญิง = 2, นาย = 3, นางสาว = 4 และค่าว่าง เป็นNA

ตารางที่2 ข้อมูลการศึกษาและการทำงานให้เปลี่ยนจากภาษาเป็นตัวเลข และค่าว่างให้เป็น NA เขียนคำสั่งในโปรแกรม R (Command R)

```
dt$titnme <- ifelse(titnme==1,"เด็กชาย",  
                    ifelse(titnme==2,"เด็กหญิง",  
                            ifelse(titnme==3,"นาย","นางสาว")))
```

```
dt$age <- ifelse(dt$age=="",NA,dt$age)
```

```
dt$gender <- ifelse(is.na(gender), titnme, gender)
```

```
dt$a9 <- ifelse(dt$a9=="ได้",1,0)
```

```
dt$a10 <- ifelse(dt$a10=="ได้",1,0)
```

```
dt$a11 <- ifelse(dt$a11 == "",NA, dt$a11)
```

```
dt$a12 <- ifelse(dt$a12 == "",NA, dt$a12)
```

```
dt$a13 <- ifelse(dt$a13 == "",NA, dt$a13)
```

```
dt$a14 <- ifelse(dt$a14 == "",NA, dt$a14)
```

2. จากข้อมูลตารางที่ 4 ข้อมูลปริมาณน้ำฝน และตารางที่ 5 ข้อมูลอุณหภูมิ ให้นักศึกษาอธิบายการผสานข้อมูล พร้อมเขียนคำสั่งในโปรแกรม R

ตารางที่ 3 ข้อมูลปริมาณน้ำฝน

ปริมาณน้ำฝน				
วันที่	ภาค	สถานีอุตุนิยมวิทยา	ปริมาณฝน 24 ชม. (มม.)	รวมตั้งแต่ต้นปี (มม.)
20120102	EE	จันทบุรี	ไม่มีฝน	0
20120104	NE	บุรีรัมย์	ไม่มีฝน	0
20120104	CC	ปทุมธานี สกษ.	ไม่มีฝน	2.8
20120105	CC	สนามบินสุวรรณภูมิ	ไม่มีฝน	5.5
20120107	EE	สตั๊ปป	ไม่มีฝน	15.3
20120108	NN	อุดรดิตถ์	ไม่มีฝน	1
20120120	SE	นราธิวาส	2	320.8
20120120	SE	ยะลา สกษ.	1.6	270
20120121	NN	อุ้มผาง	ไม่มีฝน	1.8
20120122	SW	กระบี่	2.2	328.6
20120129	NE	ขอนแก่น	ไม่มีฝน	38.7
20120129	SE	เพชรบุรี	7.8	7.8
20120130	NE	ท่าตูม	7.7	80.3
20120130	SE	พระแสง สอท.	25.2	284.9
20120131	NE	ชัยภูมิ	ไม่มีฝน	11.7
20120131	SW	ตรัง	8.8	112.7

ตารางที่ 3 ข้อมูลอุณหภูมิ

อุณหภูมิ				
วันที่	ภาค	สถานีอุตุนิยมวิทยา	อุณหภูมิ (°C) สูงสุด	อุณหภูมิ (°C) ต่ำสุด
20120104	CC	บัวชุม	31.9	23.2
20120104	CC	ปทุมธานี สกษ.	33	24.5
20120104	CC	สมุทรปราการ สกษ.	32.2	24
20120105	CC	กรุงเทพมหานคร	31.5	25.4
20120105	CC	สนามบินสุวรรณภูมิ	32.1	25.1
20120105	CC	สนามบินดอนเมือง	-	ยังไม่ได้รับรายงาน
20120106	SE	นราธิวาส	30	22
20120106	EE	สัดหีบ	32.5	22.8
20120107	NN	แม่ฮ่องสอน	29.5	16.3
20120108	NN	แพร่	31.6	18.3
20120108	NN	อุดรดิตถ์	31.5	20.5
20120120	SE	สะเดา	32.2	22.9
20120120	SE	ปัตตานี	32.2	23.7
20120120	SE	ยะลา สกษ.	33	23.6
20120120	SE	นราธิวาส	29.4	22.8
20120122	NN	แม่ฮ่องสอน	29.3	11
20120122	SW	กระบี่	33	20.5
20120122	SE	เพชรบุรี	31.1	23.3
20120131	NE	ชัยภูมิ	28.7	22.2
20120131	NE	ท่าตูม	24.5	20.5

คำอธิบายการทำความสะอาดข้อมูล และการผสานข้อมูล

ตารางที่ 4 ข้อมูลปริมาณน้ำฝน ตัวแปร ปริมาณฝน 24 ชม. (มม.) เปลี่ยนจาก ไม่มีฝน = 0

และตารางที่ 5 ข้อมูลอุณหภูมิ จัดการค่าข้อมูลจังหวัดให้อยู่ในประเภทเดียวกัน คือภาคใต้ ภาคกลาง ภาคเหนือ ภาคอีสาน และอุณหภูมิสูงสุด ไม่มี = NA อุณหภูมิต่ำสุด ที่ยังไม่ได้รับรายงาน = NA

ภาคอีสานเขียนคำสั่งในโปรแกรม R (Command R)

```
d$ปริมาณฝน 24 ชม. (มม.) <- ifelse(d$ปริมาณฝน 24 ชม. (มม.)=="ไม่มีฝน",0 , d$ปริมาณฝน 24 ชม. (มม.))
```

```
d$อุณหภูมิ (°C) สูงสุด <- ifelse(d$อุณหภูมิ (°C) สูงสุด == "",NA, d$อุณหภูมิ (°C) สูงสุด)
```

```
d$อุณหภูมิ (°C) ต่ำสุด <- ifelse(d$อุณหภูมิ (°C) ต่ำสุด== "ยังไม่ได้รับรายงาน",NA, d$อุณหภูมิ (°C) ต่ำสุด)
```