

데이터 전처리 연습문제

문제1

네 종류의 식단 A1, A2, A3, A4에 따른 식이요법이 혈액응고시간에 영향을 주는지 알고자 실험한 자료가 다음과 같다. 인자의 수준을 나타내는 변수를 `class` 로, 실험값을 `time` 으로 하여 데이터를 구성하여라.

		실험의 반복						
인자의 수준	A ₁	62	60	63	59	61		
	A ₂	62	67	71	64	65	66	
	A ₃	68	66	71	67	68	69	70
	A ₄	58	62	60	59	61	63	64

문제2

다음과 같은 두 개의 데이터를 하나로 합쳐서 엑셀 파일로 만들고자 한다. 코드를 작성하시오.

ID	READ	MATH	GRADE	ID	READ	MATH	GRADE
25	350	425	7	20	250	300	7
27	425	375	8	33	325	225	9
30	475	485	8	37	300	375	9
35	375	400	9				

문제 3

다음은 한 도시의 20개 중,고교를 대상으로 조사한 흡연율 자료이다. 도시와 농촌별에 따른 남녀 평균 흡연율을 확인할 수 있는 데이터 피벗테이블로 구성하시오.

```
https://data.hossam.kr/C02/smoke.xlsx
```

지역	구성	흡연율	지역	구성	흡연율
도시	남	0.64	농촌	여	0.56
도시	여	0.45	도시	여	0.48
농촌	공학	0.70	농촌	남	0.72
도시	남	0.85	도시	공학	0.68
농촌	남	0.72	농촌	남	0.79
도시	남	0.78	도시	남	0.90
도시	여	0.62	농촌	여	0.34
도시	남	0.79	도시	여	0.48
농촌	남	0.75	농촌	여	0.49
농촌	공학	0.81	농촌	공학	0.76

문제4

다음은 제주도 남녀 대학생의 개인정보 관리방법에 관한 설문조사 결과이다. 자료를 직접 입력하고 남녀별 메모의 종류별로 비율을 확인할 수 있는 데이터를 구성하여라

사회조사분석사 2급 2차 실기작업형 2회 기출문제와 데이터 일부분임

<https://data.hossam.kr/C02/privacy.xlsx>

학생(ID)	성별(sex)	메모 종류	학생(ID)	성별(sex)	메모 종류
1	남	3	16	여	2
2	남	2	17	여	1
3	남	2	18	여	2
4	남	1	19	여	3
5	남	2	20	여	2
6	남	3	21	여	2
7	남	2	22	여	2
8	남	2	23	여	2
9	남	1	24	여	2
10	남	1	25	여	2
11	남	2	26	여	2
12	남	3	27	여	2
13	남	2	28	여	2
14	남	1	29	여	1
15	남	2	30	여	2