

HW3 설명

[문제1]

어느 회사 직원들의 2019년, 한 해 동안의 급여 계산 관련 문제이다. 단, 개개 사원의 전월 실적이 반영되어 매달 급여가 결정되는 방식이다.

※ 원(시)자료

<q1_rawdata.txt>

※ 필요한 작업

먼저, 원시자료 <q1_rawdata.txt> 파일을 <C:\W전산통계\과제> 경로 밑에 다운받아 놓는다.

- (1) 원시자료 <rawdata.txt>를 불러들여 <one>이라는 SAS 데이터셋을 생성한다. 단, 어떤 변수(들)에 대하여는 결측인 경우도 포함되어 있다.
- (2) 위 (1)에서 생성된 <one>이라는 SAS 데이터셋을 이용하여 새로운 SAS 데이터셋 <two>를 생성하되 14개의 새로운 변수들(age, income01, income02, income03, income04, income05, income06, income07, income08, income09, income10, income11, income12, total_income)을 추가로 생성한다. 단, <two>라는 데이터셋에는 18개의 변수들(id, birth, date, dept, age, income01, income02, income03, income04, income05, income06, income07, income08, income09, income10, income11, income12, total_income)만 keep하도록 한다. 또한 변수들 birth, date, income01, income02, income03, income04, income05, income06, income07, income08, income09, income10, income11, income12, total_income 에 대하여는 아래와 같은 포맷이 적용되도록 한다.

age	입사일 당시 연령 (만연령)
income01	2019년 01월에 지급된 급여(보너스 포함) (기본 (월)급여(salary)에 전월 성과 달성에 따른 보너스 포함) ※ 2019년 01월 보너스 계산 방식은 다음과 같다. 전월 즉, 2018년 12월에 달성한 수치(y2018m12)가 매월 목표 달성 수치(goal)의 150% 이상이면 보너스는 기본 (월)급여(salary)의 200%가 지급되며, 150% 미만에서 100% 이상이면 보너스는 기본 (월)급여(salary)의 100%가 지급되며, 100% 미만이면 보너스는 지급되지 않는다.
income02	2019년 02월에 지급된 급여(보너스 포함) (기본 (월)급여(salary)에 전월 성과 달성에 따른 보너스 포함) ※ 2019년 02월 보너스 계산 방식은 다음과 같다. 전월 즉, 2019년 01월에 달성한 수치(y2019m01)가 매월 목표 달성 수치(goal)의 150% 이상이면 보너스는 기본 (월)급여(salary)의 200%가 지급되며, 150% 미만에서 100% 이상이면 보너스는 기본 (월)급여(salary)의 100%가 지급되며, 100% 미만이면 보너스는 지급되지 않는다.
.	.
.	.
.	.

income12	2019년 12월에 지급된 급여(보너스 포함) (기본 (월)급여(salary)에 전월 성과 달성에 따른 보너스 포함) ※ 2019년 12월 보너스 계산 방식은 다음과 같다. 전월 즉, 2019년 11월에 달성한 수치(y2019m11)가 매월 목표 달성 수치(goal)의 150% 이상이면 보너스는 기본 (월)급여(salary)의 200%가 지급되며, 150% 미만에서 100% 이상이면 보너스는 기본 (월)급여(salary)의 100%가 지급되며, 100% 미만이면 보너스는 지급되지 않는다.
total_income	2019년에(12개월 동안) 지급된 총 급여(보너스 포함)

birth date	포맷 (e.g) 03/05/1986 이라면 05MAR1985 으로 보이도록
income01 income02 . . . income12 total_income	포맷 (e.g) 5000000 이라면 ₩5,000,000.0 으로 보이도록 (소숫점 1자리까지)

- (3) 위 (2)에서 생성된 <two>라는 SAS 데이터셋을 이용하여 새로운 SAS 데이터셋 <final>을 생성하되 total_dept 라는 새로운 변수를 생성한다. 단, 이 <final>이라는 SAS 데이터셋에는 각 부서(dept)별로 한 개 관측치만(즉, 총 3개 관측치만) 남기고 2개 변수(dept, total_dept)만 keep 하도록 한다.

total_dept	부서(dept)별로 각 부서에 2019년에(12개월 동안) 지급된 총 급여(보너스 포함)
------------	--

total_dept	포맷 (e.g) 100000000 이라면 ₩100,000,000.0 으로 보이도록 (소숫점 1자리까지)
------------	--

- (4) 위 (2)와 (3)에서 각각 생성한 SAS 데이터셋 <two>와 <final>의 내용을 <C:\W전산통계\과제> 라는 경로 밑에 <results.pdf> 라는 별도의 pdf 파일로 출력하도록 한다. 단, <two> 라는 데이터셋 내용이 출력될 때에는 “개인별 (2019년) 보너스 및 급여” 라는 타이틀이 출력되게 하고 <final> 이라는 데이터셋 내용이 출력될 때에는 “부서별 (2019년) 총 지급된 보너스 및 급여” 라는 타이틀이 함께 출력되도록 한다. 또한 변수 total_dept 에는 “부서별 (2019년) 총 지급액” 이라는 레이블로 출력되도록 한다.

※ 위 작업들에 대한 프로그램을 작성하되, 작업 (2)에 대한 부분은 <C:\W전산통계\과제> 라는 경로 밑에 별도의 SAS 프로그램 <calculation.sas> 으로 작성을 해두고 <main.sas> 라는 SAS 프로그램 내에서 작업 (2)에 대한 부분은 불러다가 사용하도록 프로그램을 작성한다. 즉, <main.sas> 라는 SAS 프로그램에는 모든 작업((1), (2), (3), (4))에 대한 내용이 포함되어 있어야 하지만 작업 (2)에 대한 부분은 별도로 작성한 <calculation.sas>를 한 문장으로 불러다가 사용할 수 있도록 프로그램을 작성해야 한다.

[참고] <main.sas> 프로그램

```
작업 (1) 부분
.
.
.

작업 (2) 부분은 한 문장
작업 (3) 부분
.
.
.

작업 (4) 부분
.
.
.
```

[참고] <c:\W전산통계\과제> 경로 밑에 있는 <calculation.sas> 프로그램

```
/* 작업 (2) 부분 */
DATA two ..... ;
.
.
    ARRAY 문장, DO 그룹 등 이용
.
.

RUN;
```

[참고] [문제1] SAS 데이터셋 <one>

	id	birth	date	dept	salary	y2018m12	y2019m01	y2019m02	y2019m03	y2019m04	y2019m05
1	120101	9195	20162	SalesManagement	2500000	10	.	5	15	.	5
2	120102	8881	20528	Administration	2400000	15	10	.	10	10	.
3	120103	8410	19950	Sales	3000000	10	.	5	5	5	5
4	120104	10574	20315	Sales	4500000	.	.	5	10	4	.
5	120105	10916	20681	Administration	2700000	5	.	10	20	.	.
6	120106	9733	20162	Sales	2600000	20	10	.	25	10	15
7	120107	9407	20681	Sales	3600000	5	10	5	.	15	.
8	120108	10301	20681	SalesManagement	2800000	5	10	15	10	8	.

y2019m06	y2019m07	y2019m08	y2019m09	y2019m10	y2019m11	goal
5	15	10	.	5	15	10
10	5	15	10	10	.	10
.	.	20	.	5	15	15
5	10	10	5	5	5	15
.	4	10	20	5	20	10
5	15	10	.	15	15	15
10	5	8	12	5	30	15
20	15	10	.	5	10	10

[참고] [문제1] SAS 데이터셋 <two>

	id	birth	date	dept	age	income01	income02	income03	income04
1	120101	05MAR1985	15MAR2015	SalesManagement	30	₩5,000,000.0	₩2,500,000.0	₩2,500,000.0	₩7,500,000.0
2	120102	25APR1984	15MAR2016	Administration	31	₩7,200,000.0	₩4,800,000.0	₩2,400,000.0	₩4,800,000.0
3	120103	10JAN1983	15AUG2014	Sales	31	₩3,000,000.0	₩3,000,000.0	₩3,000,000.0	₩3,000,000.0
4	120104	13DEC1988	15AUG2015	Sales	26	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0
5	120105	20NOV1989	15AUG2016	Administration	26	₩2,700,000.0	₩2,700,000.0	₩5,400,000.0	₩8,100,000.0
6	120106	25AUG1986	15MAR2015	Sales	28	₩5,200,000.0	₩2,600,000.0	₩2,600,000.0	₩7,800,000.0
7	120107	03OCT1985	15AUG2016	Sales	30	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0
8	120108	15MAR1988	15AUG2016	SalesManagement	28	₩2,800,000.0	₩5,600,000.0	₩8,400,000.0	₩5,600,000.0

income05	income06	income07	income08	income09	income10	income11	income12	total_income
₩2,500,000.0	₩2,500,000.0	₩2,500,000.0	₩7,500,000.0	₩5,000,000.0	₩2,500,000.0	₩2,500,000.0	₩7,500,000.0	₩50,000,000.0
₩4,800,000.0	₩2,400,000.0	₩4,800,000.0	₩2,400,000.0	₩7,200,000.0	₩4,800,000.0	₩4,800,000.0	₩2,400,000.0	₩52,800,000.0
₩3,000,000.0	₩3,000,000.0	₩3,000,000.0	₩3,000,000.0	₩6,000,000.0	₩3,000,000.0	₩3,000,000.0	₩6,000,000.0	₩42,000,000.0
₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩54,000,000.0
₩2,700,000.0	₩2,700,000.0	₩2,700,000.0	₩2,700,000.0	₩5,400,000.0	₩8,100,000.0	₩2,700,000.0	₩8,100,000.0	₩54,000,000.0
₩2,600,000.0	₩5,200,000.0	₩2,600,000.0	₩5,200,000.0	₩2,600,000.0	₩2,600,000.0	₩5,200,000.0	₩5,200,000.0	₩49,400,000.0
₩7,200,000.0	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0	₩10,800,000.0	₩54,000,000.0
₩2,800,000.0	₩2,800,000.0	₩8,400,000.0	₩8,400,000.0	₩5,600,000.0	₩2,800,000.0	₩2,800,000.0	₩5,600,000.0	₩61,600,000.0

[참고] [문제1] SAS 데이터셋 <final>

	dept	total_dept
1	SalesManagement	₩111,600,000.0
2	Sales	₩199,400,000.0
3	Administration	₩106,800,000.0

[참고] [문제1] <C:₩전산통계₩과제> 경로 밑에 <results.pdf> 라는 파일 내용

개인별 (2019년) 보너스 및 급여

OBS	Id	birth	date	dept	age	Income01	Income02
1	120101	05MAR1985	15MAR2015	SalesManagement	30	₩5,000,000.0	₩2,500,000.0
2	120108	15MAR1988	15AUG2016	SalesManagement	28	₩2,800,000.0	₩5,600,000.0
3	120103	10JAN1983	15AUG2014	Sales	31	₩3,000,000.0	₩3,000,000.0
4	120104	13DEC1988	15AUG2015	Sales	26	₩4,500,000.0	₩4,500,000.0
5	120106	25AUG1986	15MAR2015	Sales	28	₩5,200,000.0	₩2,600,000.0
6	120107	03OCT1985	15AUG2016	Sales	30	₩3,600,000.0	₩3,600,000.0
7	120102	25APR1984	15MAR2016	Administration	31	₩7,200,000.0	₩4,800,000.0
8	120105	20NOV1989	15AUG2016	Administration	26	₩2,700,000.0	₩2,700,000.0

OBS	Income03	Income04	Income05	Income06	Income07	Income08
1	₩2,500,000.0	₩7,500,000.0	₩2,500,000.0	₩2,500,000.0	₩2,500,000.0	₩7,500,000.0
2	₩8,400,000.0	₩5,600,000.0	₩2,800,000.0	₩2,800,000.0	₩8,400,000.0	₩8,400,000.0
3	₩3,000,000.0	₩3,000,000.0	₩3,000,000.0	₩3,000,000.0	₩3,000,000.0	₩3,000,000.0
4	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0
5	₩2,600,000.0	₩7,800,000.0	₩2,600,000.0	₩5,200,000.0	₩2,600,000.0	₩5,200,000.0
6	₩3,600,000.0	₩3,600,000.0	₩7,200,000.0	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0
7	₩2,400,000.0	₩4,800,000.0	₩4,800,000.0	₩2,400,000.0	₩4,800,000.0	₩2,400,000.0
8	₩5,400,000.0	₩8,100,000.0	₩2,700,000.0	₩2,700,000.0	₩2,700,000.0	₩2,700,000.0

OBS	Income09	Income10	Income11	Income12	totalIncome
1	₩5,000,000.0	₩2,500,000.0	₩2,500,000.0	₩7,500,000.0	₩50,000,000.0
2	₩5,600,000.0	₩2,800,000.0	₩2,800,000.0	₩5,600,000.0	₩51,600,000.0
3	₩5,000,000.0	₩3,000,000.0	₩3,000,000.0	₩5,000,000.0	₩42,000,000.0
4	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩4,500,000.0	₩54,000,000.0
5	₩2,600,000.0	₩2,600,000.0	₩5,200,000.0	₩5,200,000.0	₩49,400,000.0
6	₩3,600,000.0	₩3,600,000.0	₩3,600,000.0	₩10,800,000.0	₩54,000,000.0
7	₩7,200,000.0	₩4,800,000.0	₩4,800,000.0	₩2,400,000.0	₩52,800,000.0
8	₩5,400,000.0	₩8,100,000.0	₩2,700,000.0	₩8,100,000.0	₩54,000,000.0

부서별 (2019년) 총 지급된 보너스 및 급여

OBS	dept	부서별 (2019년) 총지급액
1	SalesManagement	₩111,600,000.0
2	Sales	₩199,400,000.0
3	Administration	₩106,800,000.0

[문제2]

※ 원(시)자료

<q2_raw01.txt>, <q2_raw02.txt>

※ 필요한 작업

먼저, 원시자료 <q2_raw01.txt>과 <q2_raw02.txt> 파일을 <C:\W전산통계\과제> 경로 밑에 다운받아 놓는다.

- (1) 원시자료 <q2_raw01.txt>과 <q2_raw02.txt> 파일을 불러들여 각각 <one>과 <two> 라는 SAS 데이터셋을 생성한다. 단, 동일한 관측치들이 중복적으로 존재할 수도 있으며 어떤 변수(들)에 대하여는 결측인 경우도 포함되어 있다.
- (2) 위 (1)에서 생성된 <one>과 <two> 라는 데이터셋을 변수 no 기준으로 한 개의 SAS 데이터셋 <all>로 합한다. 단, **retain 문장을 적절히 활용**하여 동일한 관측치들이 중복적으로 존재하지 않도록 하여야 한다. 또한 기존 변수들에 추가로 **ave** 라는 변수를 생성하도록 한다.

ave	두 변수 math_per 와 eng_score 의 평균이지만 변수 eng_score 에 대하여도 math_per 와 같은 1 기준의 점수로 환산하여 계산한다. (e.g.) math_per=0.9, eng_score=80 이면 ave=(0.9+0.8)/2=0.85
-----	---

- (3) 위 (2)에서 생성된 <all>이라는 SAS 데이터셋을 이용하여 새로운 SAS 데이터셋 <final>을 생성하되 다음과 같은 3개의 새로운 변수들(f_count, m_count, rank)을 생성한다. 또한 변수 **ave**에 대하여는 아래 포맷을 적용하도록 하고 이 데이터셋 <final>에는 7개의 변수들(no, name, gender, ave, f_count, m_count, rank)만 keep되도록 한다.

f_count	gender 값이 “F”인 관측치가 한 번 나올 때마다 1씩 누적되도록 하여 이 데이터셋에서 맨 마지막 관측치에 대한 f_count 라는 변수 값을 보면 이 데이터셋에 gender 값이 “F”인 관측치가 총 몇 명인지 알 수 있음
m_count	gender 값이 “M”인 관측치가 한 번 나올 때마다 1씩 누적되도록 하여 이 데이터셋에서 맨 마지막 관측치에 대한 m_count 라는 변수 값을 보면 이 데이터셋에 gender 값이 “M”인 관측치가 총 몇 명인지 알 수 있음
rank	ave 라는 변수값 기준으로 볼 때 gender가 “F”인 그룹 내에서, 또는 “M”인 그룹 내에서 몇 등인지를 알 수 있도록 (e.g.) 이 데이터셋에 gender가 “F”인 관측치가 3개 있고 이 관측치들의 ave 값이 각각 0.6, 0.9, 0.7 이라면 ave 값이 0.9인 관측치의 rank 값은 “F 내에서 1등” ave 값이 0.7인 관측치의 rank 값은 “F 내에서 2등” ave 값이 0.6인 관측치의 rank 값은 “F 내에서 3등”

ave	포맷 (e.g) 0.878 라면 87.80% 로 보이도록 (소숫점 2자리까지)
-----	--

[참고] [문제2] SAS 데이터셋 <one>

VIEWTABLE: Work.One							
	no	math_per	eng_score	name	gender	award	
1	101	0,815	75	Jacob	M	55000	
2	101	0,815	75	Jacob	M	55000	
3	101	0,815	75	Jacob	M	55000	
4	108	0,9035	90,5	Andre	M	55000	
5	108	1	99,5	Andre	M	55000	
6	109	0,7512	100	Daniel	M	55000	
7	110	0,65	95		F	50,95	

[참고] [문제2] SAS 데이터셋 <two>

VIEWTABLE: Work.Two							
	no	math_per	eng_score	name	gender	award	
1	101	0,815	75	Jacob	M	55000	
2	101	0,815	75	Jacob	M	55000	
3	102	0,6845	80	Jenny	F	50,95	
4	103	0,905	100	Nicole	F	50,95	
5	107	0,855	70,5	Daniel	M	55000	
6	110	0,65	95		F	50,95	

[참고] [문제2] SAS 데이터셋 <all>

VIEWTABLE: Work.All							
	no	math_per	eng_score	name	gender	award	ave
1	101	0,815	75	Jacob	M	55000	0,7825
2	102	0,6845	80	Jenny	F	50,95	0,74225
3	103	0,905	100	Nicole	F	50,95	0,9525
4	107	0,855	70,5	Daniel	M	55000	0,78
5	108	0,9035	90,5	Andre	M	55000	0,90425
6	109	0,7512	100	Daniel	M	55000	0,8756
7	110	0,65	95		F	50,95	0,8

[참고] [문제2] SAS 데이터셋 <final>

VIEWTABLE: Work.Final							
	no	name	gender	ave	f_count	m_count	rank
1	103	Nicole	F	95,25%	1	0	F 내에서 1등
2	110		F	80,00%	2	0	F 내에서 2등
3	102	Jenny	F	74,23%	3	0	F 내에서 3등
4	108	Andre	M	90,43%	3	1	M 내에서 1등
5	109	Daniel	M	87,56%	3	2	M 내에서 2등
6	101	Jacob	M	78,25%	3	3	M 내에서 3등
7	107	Daniel	M	78,00%	3	4	M 내에서 4등

▶ **과제 제출방법: [문제1]과 [문제2] 각각에 대한 프로그램 파일**

[문제1]에 대한 SAS 프로그램 파일 2개(main.sas, calculation.sas)와 [문제2]에 대한 SAS 프로그램 파일 1개 제출 (**총 3개 파일**)

제 출 일: 6/21(일) 오후 11시까지

제출방법: 스노우보드 <HW3>에 파일(SAS 프로그램 파일 3개) 업로드