

SAS 실습과제 2

1. sashelp.heart 데이터셋을 가져와서, 다음 2 개의 변수(BMI, BMI_Q)를 새롭게 생성하여 추가한 뒤, BMI_Q 변수값이 “ Q4 ” 인 observation 만 포함하는 새로운 데이터셋을 생성하여 work.heart3 라는 데이터셋으로 저장하는 SAS code 를 작성하시오.

- **BMI** : 체질량지수 (kg/m^2 단위)
- **BMI_Q**: BMI 변수값을 Quartile in category 로 나누어 다음과 같은 값(categorical data)을 부여

BMI 변수값	BMI_Q 변수값
누적백분율 $0\% \leq \text{BMI} < \text{누적백분율}25\%$	Q1
누적백분율 $25\% \leq \text{BMI} < \text{누적백분율}50\%$	Q2
누적백분율 $50\% \leq \text{BMI} < \text{누적백분율}75\%$	Q3
누적백분율 $75\% \leq \text{BMI} \leq \text{누적백분율}100\%$	Q4

- 참고로 sashelp.heart 데이터셋의 Height 변수값의 단위는 ‘ 인치(in) ’, Weight 변수값의 단위는 ‘ 파운드(lb) ’ 임.
- Work.heart3 데이터셋이 최종적으로 포함할 변수(총 12 개): Status, AgeCHDdiag, AgeAtStart, AgeAtDeath, Sex, Height, Weight, Systolic, Cholesterol, Smoking_Status, BMI, BMI_

```
[SAS code] data work.heart;
            set sashelp.heart;
            run;
            data work.heart2;
            set work.heart;
            weight_kg= weight*0.453592;
            height_m= height*0.0254;
            BMI=weight_kg/(height_m**2);
            run;

            proc univariate data=work.heart2 ;
            var BMI;
            run;
            proc univariate data=work.heart2 noprint;
            var BMI;
            output out = work.h5
            pctlpts = 0 25 50 75 100
            pctlpre = Quint;
            run;
            data work.heart3;
            set work.heart2;
            if 14.1243<=BMI<22.6551 then BMI_Q="Q1";
            else if 22.6551<=BMI<25.1298 then BMI_Q="Q2";
            else if 25.1298<=BMI<27.8911 then BMI_Q="Q3";
            else if 27.8911<=BMI<=56.6838 then BMI_Q="Q4";
            run;
            data work.heart3;
            set work.heart3;
            where BMI_Q="Q4";
            keep Status AgeCHDdiag AgeAtStart AgeAtDeath Sex Height Weight Systolic Cholesterol
            Smoking_Status BMI BMI_Q;
            run;
```

2. 위에서 생성한 work.heart3 데이터셋을 이용하여 다음의 표를 완성하시오.

(회색 글씨는 표시 형식을 보여주는 예이므로 삭제하고, 실제 숫자를 확인후 검은색 글씨로 채워 넣으시오)

[PROC MEANS, PROC FREQ 이용]

Table 1. Baseline characteristics of study population

	Men	Women
n	599	700
Age at baseline, years	44.559 ± 8.6	47.941 ± 8.1
Body Mass Index, kg/m ²	30.31 ± 2.221	31.8489 ± 3.9
Systolic blood pressure, mmHg	144.38 ± 22.448	151.37 ± 29.569
Total cholesterol, mg/dL	231.1 ± 44.282	237.63 ± 46.491
Non-smoker, n (%)	262(44.18)	505 (72.56)

```
[SAS code] proc means data=work.heart3 ;
class sex;
run;
proc freq data=work.heart3;
table Smoking_Status*sex ;
run;
```

3. sashelp.heart 데이터셋을 이용하여 남자와 여자의 체질량지수(kg/m² 단위)를 각각 히스토그램으로 그려서 제시하시오. [그림 2 개와 필요한 SAS code 를 제출할 것]

그림 1. 체질량지수 히스토그램 (남자 599명)

그림 2. 체질량지수 히스토그램 (여자 700명)

```
[SAS code] /*문제가 혼동되어서 두 코드 모두 첨부합니다*/
/*work.heart3사용해서 히스토그램*/
proc univariate data=work.heart3;
class sex;
histogram BMI;
run;
/*sashelp.heart사용해서 히스토그램그릴 경우*/
data work.heart4;
set sashelp.heart;
weight_kg= weight*0.453592;
height_m= height*0.0254;
BMI=weight_kg/(height_m**2);
run;
proc univariate data=work.heart4;
class sex;
histogram BMI;
run;
```

