

SAS 실습과제 3

t-Test / Chi-square Test / ANOVA

1. 중간고사 6.(1) 문제에서 작성한 work.midterm 데이터셋을 이용하여 중간고사 6.(2)의 Table 1에 대하여 t-Test 및 Chi-square Test를 실시한 뒤 P value를 구하여 다음의 Table 1-1을 완성하시오.

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

* 'UI/Cre'와 'TPOAb'는 중간고사 문제 6.(2)와 다르게 'Mean±SD'를 제시하는 것으로 바뀌었음 (나머지 항목은 기존의 숫자를 그대로 활용 가능)

* 결측치에 대해서는 무시하고 SAS 에서 출력되는 Mean±SD, n(%), P value 를 그대로 채워 넣을 것

Table 1-1. Baseline characteristics of study population

	Men	Women	P value
n	1060	935	
Age, years	37.272 ± 15.929	38.952 ± 15.916	0.0188
Body Mass Index, kg/m ²	23.871 ± 3.609	22.713 ± 3.508	<.0001
Current smoker, n (%)	363(22.7)	42 (2.63)	<.0001
Ever drinker, n (%)	874 (45.05)	710 (36.6)	0.5781
UI/Cre, µg/g	4.748 ± 11.359	6.901 ± 17.148	0.0008
Blood cadmium, µg/L	0.804 ± 0.681	0.959 ± 0.643	<.0001
TSH, µIU/mL	2.561 ± 2.563	2.905 ± 2.187	0.0014
ft4, ng/dL	1.277 ± 0.179	1.19 ± 0.266	<.0001
TPOAb, IU/mL	12.835 ± 47.212	34.674 ± 162.783	<.0001

```
libname mid '/home/u63345097/data for midterm';
data work.midterm;
  set mid.HN13_ALL(encoding=any);
  where HE_Cd ne .;
```

[SAS code]

```
if age > 65 or LW_mp = 1 or DC7_dg = 1 or TH_tx = 1 or HE_Ulod = . or HE_UCREA = . then delete;
```

```
UICre = HE_Ulod / HE_UCREA;
```

```
keep age sex HE_BMI HE_Cd HE_TSH HE_ft4 HE_TPOAb HE_Ulod HE_UCREA UICre sm_presnt BD1;
```

```
run;
```

```
proc means data=work.midterm mean std;
```

```
  class sex; /* 성별에 따라 분할 */
```

```
  var HE_BMI HE_Cd HE_TSH HE_ft4 HE_TPOAb HE_Ulod HE_UCREA UICre;
```

```
run;
```

```
proc ttest data=work.midterm;
```

```
  class sex;
```

```
  var age ;
```

```
run;
```

```
PROC FREQ data=work.midterm;
```

```
  TABLE sex*sm_presnt/CHISQ;
```

```
RUN;
```

```
PROC FREQ data=work.midterm;
```

```
  TABLE sex*BD1/CHISQ;
```

```
RUN;
```

2. 중간고사 6.(1) 문제에서 작성한 work.midterm 데이터셋을 가져와서, 다음의 변수(HE_Cd_Q)를 새롭게 생성하여 추가한 뒤 work.quiz3 라는 데이터셋으로 저장하는 SAS code 를 작성하시오.

- **HE_Cd_Q**: HE_Cd 변수값을 Quartile in category 로 나누어 다음과 같은 값(categorical data)을 부여

HE_Cd 변수값	BMI_Q 변수값
누적백분율 0% ≤ HE_Cd < 누적백분율 25%	1
누적백분율 25% ≤ HE_Cd < 누적백분율 50%	2
누적백분율 50% ≤ HE_Cd < 누적백분율 75%	3
누적백분율 75% ≤ HE_Cd ≤ 누적백분율 100%	4

[SAS code]

```
proc univariate data=work.midterm ; data work.quiz3;
var HE_Cd; set work.midterm;
run; if 0.021 <= HE_Cd < 0.415 then HE_Cd_Q = 1;
else if 0.42 <= HE_Cd < 0.76 then HE_Cd_Q = 2;
else if 0.76 <= HE_Cd < 1.186 then HE_Cd_Q = 3;
else if 1.186 <= HE_Cd <= 11.046 then HE_Cd_Q = 4;
run;
```

3. 위에서 생성한 work.quiz3 데이터셋을 이용하여 ANOVA 검정을 실시하고 다음의 Table 2 를 완성하시오.

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

Table 2. Thyroid hormone levels according to the quartiles of blood cadmium levels

	BCd Q1	BCd Q2	BCd Q3	BCd Q4	P value
n	497	499	499	500	
Blood cadmium, µg/L	0.021 – <0.42	0.415– <0.76	0.76 <1.186	1.186– 11.046	
TSH, µIU/mL	2.701± 1.669	2.664± 1.788	2.826 ± 3.437	2.698 ± 2.293	0.7241
ft4, ng/dL	1.265 ± 0.182	1.242 ± 0.197	1.223 ± 0.200	1.215 ± 0.308	0.0027
TPOAb, IU/mL	14.35 ± 67.141	25.099 ± 127.68	23.171 ± 104.289	29.612 ± 151.906	0.2154

BCd, blood cadmium; TSH, thyroid-stimulating hormone; ft4, free thyroxine; TPOAb, anti-thyroid peroxidase antibody.

[SAS code]

```
proc freq data=work.quiz3;
table HE_Cd_Q ;
run;
proc glm data=work.quiz3;
class HE_Cd_Q;
model HE_TSH HE_ft4 HE_TPOAb = HE_Cd_Q;
means HE_Cd_Q;

run;
```