

이름: 김연주

학과: 보건정책관리학부

학번: 2021250461

국민건강조사자료 샘플데이터(knhanes.csv)를 사용하여 다음의 분석을 시행하시오. (전체 n= 8127)

1. 다음의 변수에 대하여 코드북(3-4쪽)을 확인하고 결측값이 '8' 또는 '9'로 표기된 경우에 한하여 결측값(NA)으로 인식되도록 처리하시오.

대상변수: 고혈압 유병 여부(DI1_pr), 성별(sex), 현재 흡연 여부(BS3_1), 평생음주경험(BD1), 결혼여부(Marri_1), 주관적 건강인지(D_1_1), 이상지질증 여부(DI2_pr), 당뇨병 여부(DE1_pr)

※ 주의: 현재 흡연 여부(BS3_1) 변수의 경우 '8' 은 비흡연자를 나타내므로 결측으로 처리하지 않고 '9' 만 결측값으로 처리한다.

script:

```

8
9 #1번
10
11 ##import dataset 통해 knhanes.csv 불러오기
12
13 ##ifelse 문을 사용하여 결측값 처리
14 ###고혈압 현재 유병 여부: 8,9 결측값 처리
15 knhanes$DI1_pr_n=as.factor(ifelse(knhanes$DI1_pr<8, knhanes$DI1_pr, NA))
16 ###현재흡연 여부: 9 결측값 처리
17 knhanes$BS3_1_n=as.factor(ifelse(knhanes$BS3_1<9, knhanes$BS3_1, NA))
18 ###평생음주경험: 8,9 결측값 처리
19 knhanes$BD1_n=as.factor(ifelse(knhanes$BD1<8, knhanes$BD1, NA))
20 ###결혼 여부: 9 결측값 처리
21 knhanes$marri_1_n=as.factor(ifelse(knhanes$marri_1<9, knhanes$marri_1, NA))
22 ###주관적 건강인지: 9 결측값 처리
23 knhanes$D_1_1_n=as.factor(ifelse(knhanes$D_1_1<9, knhanes$D_1_1, NA))
24 ###이상지질혈증 현재 유병 여부: 8,9 결측값 처리
25 knhanes$DI2_pr_n=as.factor(ifelse(knhanes$DI2_pr<8, knhanes$DI2_pr, NA))
26 ###당뇨병 현재 유병 여부: 8,9 결측값 처리
27 knhanes$DE1_pr_n=as.factor(ifelse(knhanes$DE1_pr<8, knhanes$DE1_pr, NA))
28
29

```

console:

```
> ##ifelse 문을 사용하여 결측값 처리
> ###고혈압 현재 유병 여부: 8,9 결측값 처리
> knhanes$DI1_pr_n=as.factor(ifelse(knhanes$DI1_pr<8, knhanes$DI1_pr, NA))
> ###현재흡연 여부: 9 결측값 처리
> knhanes$BS3_1_n=as.factor(ifelse(knhanes$BS3_1<9, knhanes$BS3_1, NA))
> ###평생음주경험: 8,9 결측값 처리
> knhanes$BD1_n=as.factor(ifelse(knhanes$BD1<8, knhanes$BD1, NA))
> ###결혼 여부: 9 결측값 처리
> knhanes$marri_1_n=as.factor(ifelse(knhanes$marri_1<9, knhanes$marri_1, NA))
> ###주관적 건강인지: 9 결측값 처리
> knhanes$D_1_1_n=as.factor(ifelse(knhanes$D_1_1<9, knhanes$D_1_1, NA))
> ###이상지질혈증 현재 유병 여부: 8,9 결측값 처리
> knhanes$DI2_pr_n=as.factor(ifelse(knhanes$DI2_pr<8, knhanes$DI2_pr, NA))
> ###당뇨병 현재 유병 여부: 8,9 결측값 처리
> knhanes$DE1_pr_n=as.factor(ifelse(knhanes$DE1_pr<8, knhanes$DE1_pr, NA))
> |
```

설명:

"ifelse"를 사용하여 결측값을 NA로 인식되도록 처리하였다. 고혈압 현재 유병 여부 (DI1_pr), 평생음주경험 (BD1), 이상지질혈증 현재 유병 여부 (DI2_pr), 당뇨병 현재 유병 여부 (DE1_pr)은 8, 9를 결측값으로 인식하도록 처리하였고, 현재흡연 여부(BS3_1), 결혼 여부 (marri_1), 주관적 건강인지 (D_1_1)은 9를 결측값으로 인식하도록 처리하였다. 처리한 변수들은 변수명 뒤에 '_n'을 붙여 정의하였다.

2. 연속변수인 수축기 혈압(HE_sbp), 연령(age), 신장(HE_ht), 체중(HE_wt), 허리둘레(HE_wc), 공복혈당(HE_glu),

총콜레스테롤(HE_chol), EQ-5D 지표(EQ5D)의 기술통계(n, 평균, 표준편차, 최대, 최소)를 구하시오.

Script:

```
34 #2번
35
36 ##psych 패키지 설치 후 사용 준비
37 install.packages("psych")
38 library(psych)
39
40 ##describe로 기술통계 요청
41 ###최종 수축기 혈압
42 describe(knhanes$HE_sbp)
43 ###만나이
44 describe(knhanes$age)
45 ###신장
46 describe(knhanes$HE_ht)
47 ###체중
48 describe(knhanes$HE_wt)
49 ###허리둘레
50 describe(knhanes$HE_wc)
51 ###공복혈당
52 describe(knhanes$HE_glu)
53 ###총콜레스테롤
54 describe(knhanes$HE_chol)
55 ###EQ-5D index
56 describe(knhanes$EQ5D)
57
```

Console:

```
The downloaded source packages are in
      'C:\Users\mkjhoe\AppData\Local\Temp\RtmpC4ouhR\downloaded_packages'
> library(psych)
> ##describe로 기술통계 요청
> ###최종 수축기 혈압
> describe(knhanes$HE_sbp)
  vars    n mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 6878 117.97 16.68   116 116.73 16.31  77 232   155 0.94     1.95 0.2
> ###만나이
> describe(knhanes$age)
  vars    n mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 8127  43.12 22.73    45  43.58 26.69   1  80    79 -0.2     -1.06 0.25
> ###신장
> describe(knhanes$HE_ht)
  vars    n mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 7684 157.42 18.65  160.8 160.48 11.42 73.9 194 120.1 -1.91     4.34 0.21
> ###체중
> describe(knhanes$HE_wt)
  vars    n mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 7695  58.37 17.79   59.5  59.55 13.49  7.9 140 132.1 -0.49     0.92 0.2
> ###허리둘레
> describe(knhanes$HE_wc)
  vars    n mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 7662  77.7 13.79  79.05  78.4 13.12 24.5 128.3 103.8 -0.37     0.09 0.16
```

```

> ###공복혈당
> describe(knhanes$HE_glu)
  vars   n  mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 6545 100.41 24.01     95   96.23 10.38 54 379   325 4.39   29.63 0.3
> ###총콜레스테롤
> describe(knhanes$HE_chol)
  vars   n  mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 6546 190.4 38.09    188  189.04 37.06 64 451   387 0.42    0.56 0.47
> ###EQ-5D index
> describe(knhanes$EQ5D)
  vars   n  mean    sd median trimmed mad  min max range  skew kurtosis se
X1     1 5833 0.95 0.11     1    0.97  0 -0.17  1  1.17 -3.47   17.99  0
>

```

설명:

기술 통계를 구하기 위해 'psych' 패키지를 설치 및 실행하였다. 이후 'describe' 문을 사용하여 기술통계를 구하였다. 수축기혈압 (HE_sbp), 연령 (age), 신장 (HE_ht), 체중 (HE_wt), 허리둘레 (HE_wc), 공복혈당 (HE_glu), 총콜레스테롤 (HE_chol), EQ-5D 지표 (EQ5D)는 연속변수이므로 n, 평균, 표준편차, 최대, 최소를 구할 수 있다.

Console 사진에서 확인할 수 있는 n 은 n, mean 은 평균, sd 는 표준편차, min 은 최소, max 는 최대값이다. 이를 통해 아래와 같은 기술통계 결과를 알 수 있다.

	n	평균	표준편차	최대	최소
HE_sbp	6878	117.97	16.68	232	77
Age	8127	43.12	22.73	80	1
HE_ht	7684	157.42	18.65	194	73.9
HE_wt	7695	58.37	17.79	140	7.9
HE_wc	7662	77.7	13.79	128.3	24.5
HE_glu	6545	100.41	24.01	379	54
HE_chol	6546	190.4	38.09	451	64
EQ5D	5833	0.95	0.11	1	-0.17

3. 범주형 변수인 고혈압 유병 여부(DI1_pr), 성별(sex), 현재 흡연 여부(BS3_1), 평생음주경험(BD1), 결혼여부(Marri_1), 소득 5분위수(ho_incm5), 교육수준(edu), 주관적 건강인지(D_1_1), 이상지질증 여부(DI2_pr), 당뇨병 여부(DE1_pr)의 범주별 빈도와 퍼센트를 구하시오.

Script

```
62 #3번
63
64 ##prettyR 패키지 설치, 사용 준비
65 install.packages("prettyR")
66 library(prettyR)
67
68 ##freq 사용하여 범주형변수 빈도, 퍼센트 요청
69
70 ###고혈압 현재 유병 여부 (결측값 처리 이전)
71 freq(knhanes$DI1_pr)
72 ###고혈압 현재 유병 여부 (결측값 처리 이후)
73 freq(knhanes$DI1_pr_n)
74 ###성별
75 freq(knhanes$sex)
76 ###현재흡연 여부 (결측값 처리 이전)
77 freq(knhanes$BS3_1)
78 ###현재흡연 여부 (결측값 처리 이후)
79 freq(knhanes$BS3_1_n)
80 ###평생음주경험 (결측값 처리 이전)
81 freq(knhanes$BD1)
82 ###평생음주경험 (결측값 처리 이후)
83 freq(knhanes$BD1_n)
84 ###결혼여부 (결측값 처리 이전)
85 freq(knhanes$marri_1)
86 ###결혼여부 (결측값 처리 이후)
87 freq(knhanes$marri_1_n)
88 ###소득 5분위수 (가구)
89 freq(knhanes$ho_incm5)
90 ###교육수준 재분류 코드
91 freq(knhanes$edu)
92 ###주관적 건강 인지 (결측값 처리 이전)
93 freq(knhanes$D_1_1)
94 ###주관적 건강 인지 (결측값 처리 이후)
95 freq(knhanes$D_1_1_n)
96 ###이상지질혈증 현재 유병 여부 (결측값 처리 이전)
97 freq(knhanes$DI2_pr)
98 ###이상지질혈증 현재 유병 여부 (결측값 처리 이후)
99 freq(knhanes$DI2_pr_n)
100 ###당뇨병 현재 유병 여부 (결측값 처리 이전)
101 freq(knhanes$DE1_pr)
102 ###당뇨병 현재 유병 여부 (결측값 처리 이후)
103 freq(knhanes$DE1_pr_n)
104
```

Console:

```
The downloaded binary packages are in
  C:\Users\Wkjhoe\AppData\Local\Temp\RtmpC4ouhR\downloaded_packages
> library(prettyR)
```

다음의 패키지를 부착합니다: 'prettyR'

The following objects are masked from 'package:psych' :

describe, skew

```
> ###고혈압 현재 유병 여부 (결측값 처리 이전)
> freq(knhanes$DI1_pr)
```

```
Frequencies for knhanes$DI1_pr
      8      1      0      9      NA
6169 1449    79     2   428
%    75.9 17.8    1    0   5.3
%!NA 80.1 18.8    1    0
```

```
> ###고혈압 현재 유병 여부 (결측값 처리 이후)
> freq(knhanes$DI1_pr_n)
```

```
Frequencies for knhanes$DI1_pr_n
      1      0      NA
1449    79 6599
%    17.8    1 81.2
%!NA 94.8  5.2
```

```
> ###성별
> freq(knhanes$sex)
```

```
Frequencies for knhanes$sex
      2      1      NA
4361 3766     0
%    53.7 46.3    0
%!NA 53.7 46.3
```

```
> ###현재흡연 여부 (결측값 처리 이전)
> freq(knhanes$BS3_1)
```

```
Frequencies for knhanes$BS3_1
      8      3      1      2      9      NA
5199 1331  941  153   75  428
%      64 16.4 11.6  1.9  0.9  5.3
%!NA 67.5 17.3 12.2   2   1
```

```
> ###현재흡연 여부 (결측값 처리 이후)
> freq(knhanes$BS3_1_n)
```

```
Frequencies for knhanes$BS3_1_n
      8      3      1      2      NA
5199 1331  941  153  503
%      64 16.4 11.6  1.9  6.2
%!NA 68.2 17.5 12.3   2
```

```
> ###평생음주경험 (결측값 처리 이전)
> freq(knhanes$BD1)
```

```
Frequencies for knhanes$BD1
      2      1      8      9      NA
5561 1083  990   65  428
%      68.4 13.3 12.2  0.8  5.3
%!NA 72.2 14.1 12.9  0.8
```

```
> ###평생음주경험 (결측값 처리 이후)
> freq(knhanes$BD1_n)
```

```
Frequencies for knhanes$BD1_n
      2      1      NA
5561 1083 1483
%      68.4 13.3 18.2
%!NA 83.7 16.3
```

```
> ###결혼여부 (결측값 처리 이전)
> freq(knhanes$marri_1)
```

```
Frequencies for knhanes$marri_1
      1      2      NA
5386 2740   1
%      66.3 33.7   0
%!NA 66.3 33.7
```

```
> ###결혼여부 (결측값 처리 이후)
```

```
> freq(knhanes$marri_1_n)
```

```
Frequencies for knhanes$marri_1_n
```

	1	2	NA
	5386	2740	1
%	66.3	33.7	0
%!NA	66.3	33.7	

```
> ###소득 5분위수 (가구)
```

```
> freq(knhanes$ho_incm5)
```

```
Frequencies for knhanes$ho_incm5
```

	5	3	4	2	1	NA
	1900	1842	1830	1431	1096	28
%	23.4	22.7	22.5	17.6	13.5	0.3
%!NA	23.5	22.7	22.6	17.7	13.5	

```
> ###교육수준 재분류 코드
```

```
> freq(knhanes$edu)
```

```
Frequencies for knhanes$edu
```

	1	4	3	2	NA
	2501	2212	1812	804	798
%	30.8	27.2	22.3	9.9	9.8
%!NA	34.1	30.2	24.7	11	

```
> ###주관적 건강 인지 (결측값 처리 이전)
```

```
> freq(knhanes$D_1_1)
```

```
Frequencies for knhanes$D_1_1
```

	3	2	4	1	9	5	NA
	3565	1996	975	590	334	239	428
%	43.9	24.6	12	7.3	4.1	2.9	5.3
%!NA	46.3	25.9	12.7	7.7	4.3	3.1	

```
> ###주관적 건강 인지 (결측값 처리 이후)
```

```
> freq(knhanes$D_1_1_n)
```

```
Frequencies for knhanes$D_1_1_n
```

	3	2	4	1	5	NA
	3565	1996	975	590	239	762
%	43.9	24.6	12	7.3	2.9	9.4
%!NA	48.4	27.1	13.2	8	3.2	


```
> ###이상지질혈증 현재 유병 여부 (결측값 처리 이전)
```

```
> freq(knhanes$DI2_pr)
```

```
Frequencies for knhanes$DI2_pr
```

	8	1	0	9	NA
	6530	940	227	2	428
%	80.3	11.6	2.8	0	5.3
%!NA	84.8	12.2	2.9	0	

```
> ###이상지질혈증 현재 유병 여부 (결측값 처리 이후)
```

```
> freq(knhanes$DI2_pr_n)
```

```
Frequencies for knhanes$DI2_pr_n
```

	1	0	NA
	940	227	6960
%	11.6	2.8	85.6
%!NA	80.5	19.5	

```
> ###당뇨병 현재 유병 여부 (결측값 처리 이전)
```

```
> freq(knhanes$DE1_pr)
```

```
Frequencies for knhanes$DE1_pr
```

	8	1	0	9	NA
	7100	579	16	4	428
%	87.4	7.1	0.2	0	5.3
%!NA	92.2	7.5	0.2	0.1	

```
> ###당뇨병 현재 유병 여부 (결측값 처리 이후)
```

```
> freq(knhanes$DE1_pr_n)
```

```
Frequencies for knhanes$DE1_pr_n
```

	1	0	NA
	579	16	7532
%	7.1	0.2	92.7
%!NA	97.3	2.7	

설명:

고혈압 유병 여부 (DI1_pr), 성별 (sex), 현재 흡연 여부 (BS3_1), 평생음주경험 (BD1), 결혼 여부 (marri_1), 소득 5 분위수 (ho_incm5), 교육수준 (edu), 주관적 건강 인지 (D_1_1), 이상 지질증 여부 (DI2_pr), 당뇨병 여부 (DE1_pr)은 범주형 변수이므로 'freq'문을 통해 빈도와 퍼센트를 구할 수 있다. 'freq'를 사용하기 위해 통계 패키지인 'prettyR'을 설치 및 실행하였다.

또한, 이중 고혈압 유병 여부 (DI1_pr), 현재 흡연 여부 (BS3_1), 평생음주경험 (BD1), 결혼 여부 (marri_1), 주관적 건강 인지 (D_1_1), 이상 지질증 여부 (DI2_pr), 당뇨병 여부 (DE1_pr)은 위에서 결측값 처리를 해주었던 변수들이므로 결측값 처리 전후의 빈도와 퍼센트를 따로 나타낼 수 있다. 이러한 결과를 console 창에서 확인

가능하였으며, 해당 결과를 표로 정리하면 아래와 같다. '결측 이전 빈도', '결측 이전 %', '결측 이전 %!NA'는 각각 결측값 처리 이전 빈도, 퍼센트, %!NA 를 나타내고, '결측 이후 빈도', '결측 이후 %', '결측 이후 %!NA'는 각각 결측값 처리 이후 빈도, 퍼센트, %!NA 를 나타낸다.

DI1_pr	0	1	8	9	NA
결측 이전 빈도	79	1449	6169	2	428
결측 이전 %	1	17.8	75.9	0	5.3
결측 이전 %!NA	1	18.8	80.1	0	
결측 이후 빈도	79	1449			6599
결측 이후 %	1	17.8			81.2
결측 이후 %!NA	5.2	94.8			

sex	1	2	NA
빈도	3766	4361	0
%	46.3	53.7	0
%!NA	46.3	53.7	

BS3_1	1	2	3	8	9	NA
결측 이전 빈도	941	153	1331	5199	75	428
결측 이전 %	11.6	1.9	16.4	64	0.9	5.3
결측 이전 %!NA	12.2	2	17.3	67.5	1	
결측 이후 빈도	941	153	1331	5199		503
결측 이후 %	11.6	1.9	16.4	64		6.2
결측 이후 %!NA	12.3	2	17.5	68.2		

BD1	1	2	8	9	NA
결측 이전 빈도	1083	5561	990	65	428
결측 이전 %	13.3	68.4	12.2	0.8	5.3
결측 이전 %!NA	14.1	72.2	12.9	0.8	
결측 이후 빈도	1083	5561			1483
결측 이후 %	13.3	68.4			18.2
결측 이후 %!NA	16.3	83.7			

marri_1	1	2	NA
결측 이전 빈도	5386	2740	1
결측 이전 %	66.3	33.7	0
결측 이전 %!NA	66.3	33.7	
결측 이후 빈도	5386	2740	1
결측 이후 %	66.3	33.7	0
결측 이후 %!NA	66.3	33.7	

결혼 여부 변수는 9로 응답한 관측치가 존재하지 않기 때문에 결측값 처리 전후 빈도가 같다.

ho_incm5	1	2	3	4	5	NA
빈도	1096	1431	1842	1830	1900	28
%	13.5	17.6	22.7	22.5	23.4	0.3
%!NA	13.5	17.7	22.7	22.6	23.5	

edu	1	2	3	4	NA
빈도	2501	804	1812	2212	798
%	30.8	9.9	22.3	27.2	9.8
%!NA	34.1	11	24.7	30.2	

D_1_1	1	2	3	4	5	9	NA
결측 이전 빈도	590	1996	3565	975	239	334	428
결측 이전 %	7.3	24.6	43.9	12	2.9	4.1	5.3
결측 이전 %!NA	7.7	25.9	46.3	12.7	3.1	4.3	
결측 이후 빈도	590	1996	3565	975	239		762
결측 이후 %	7.3	24.6	43.9	12	2.9		9.4
결측 이후 %!NA	8	27.1	48.4	13.2	3.2		

DI2_pr	0	1	8	9	NA
결측 이전 빈도	227	940	6530	2	428
결측 이전 %	2.8	11.6	80.3	0	5.3
결측 이전 %!NA	2.9	12.2	84.8	0	
결측 이후 빈도	227	940			6960
결측 이후 %	2.8	11.6			85.6
결측 이후 %!NA	19.5	80.5			

DE1_pr	0	1	8	9	NA
결측 이전 빈도	16	579	7100	4	428
결측 이전 %	0.2	7.1	87.4	0	5.3
결측 이전 %!NA	0.2	7.5	92.2	0.1	
결측 이후 빈도	16	579	7100	4	7532
결측 이후 %	0.2	7.1	87.4	0	92.7
결측 이후 %!NA	2.7	97.3			

4. 공복혈당(HE_glu)의 값이 성별(sex)에 따라 다른지에 대하여 분석하고자 한다.

1) 남녀별로 공복혈당의 기술통계를 구하여 비교하고 상자그림을 그려 분포에 대하여 비교 설명하시오.

Script:

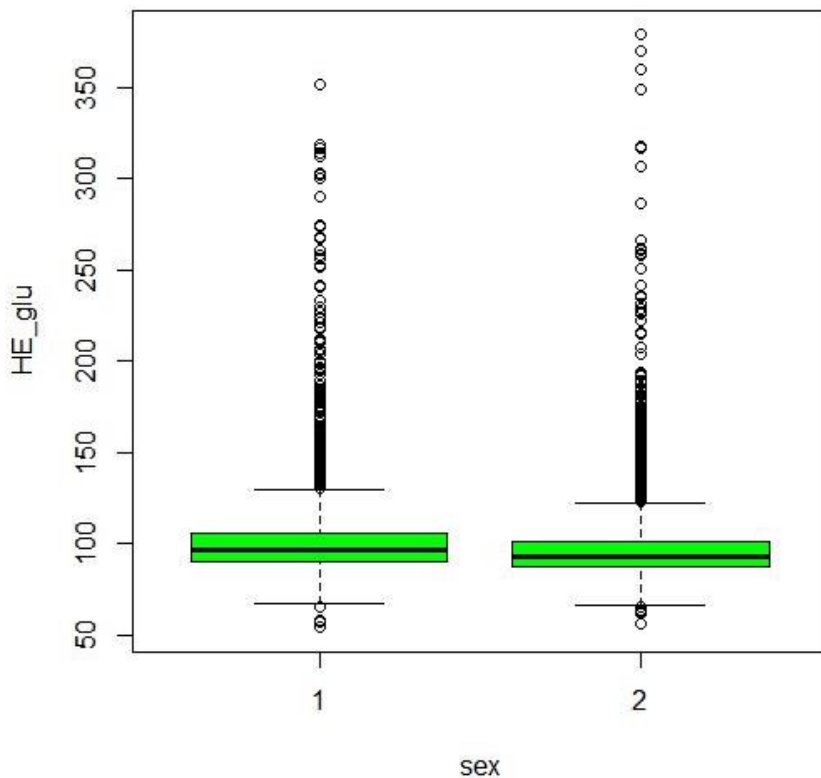
```
119 #4번
120
121 #4-1)
122
123 #psych 패키지 다시 실행
124 library(psych)
125 |
126 #describeBy를 통해 sex에 따른 HE_glu 구하기
127 describeBy(knhanes$HE_glu, knhanes$sex)
128
129 #boxplot 이용하여 상자그림 그리기
130 boxplot(HE_glu~sex, data=knhanes, col="green")
131
132
```

Console:

```
> # describeBy를 통해 sex에 따른 HE_glu 구하기
> describeBy(knhanes$HE_glu, knhanes$sex)

Descriptive statistics by group
group: 1
  vars    n  mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 2979 103.02 25.49     97   98.55 11.86  54 352   298  3.95    22.72 0.47
-----
group: 2
  vars    n  mean    sd median trimmed  mad min max range skew kurtosis  se
X1     1 3566  98.24 22.46     93   94.36  8.9   56 379   323  4.91    39.01 0.38
#> Warning: boxplot() data not sorted; sorting data to match boxplot()
#> Warning: boxplot() data not sorted; sorting data to match boxplot()
```


Box plot:



설명:

남녀별로 공복혈당의 기술통계를 구하기 위해 'psych'패키지 실행 후 'describeBy'를 활용하였다. 그 결과 아래표와 같은 성별에 따른 기술통계를 구하였다.

	n	평균	표준편차	중앙값	최소	최대	왜도	첨도
남성	2979	103.02	25.49	97	54	352	3.95	22.72
여성	3566	98.24	22.46	93	56	379	4.91	39.01

기술 통계 결과를 분석해보면 남성의 평균이 여성의 평균보다 높다. 또한, 남성의 표준편차가 비교적 큰 점을 통해 남성 공복혈당 관찰치가 여성에 비해 고르게 분포되어 있음을 알 수 있다. 이러한 분포는 왜도와 첨도를 통해서도 예측할 수 있다. 여성은 남성에 비해 왜도와 첨도가 모두 높은데, 이는 여성 공복혈당 분포가 오른쪽으로 치우쳐진 정도가 비교적 크고, 최빈값을 갖는 값이 많다는 점을 알 수 있다.

Box plot을 보면 남성의 IQR이 여성의 것보다 크다는 것과 중앙값이 크다는 점을 발견할 수 있다. 또한, 여성의 box plot을 보면 극단값 경계보다 큰 극단치가 비교적 길게 형성되어 있는데, 이를 통해 여성 공복혈당 분포의 오른쪽 꼬리가 길다는 점을 알 수 있다.

2) 남녀 집단의 공복혈당의 모분산이 같은지 검정하시오. ($\alpha=0.05$)

Script:

```
172  
173 #4-2)  
174  
175 #성별에 따른 공복혈당 등분산 검정  
176 var.test(HE_glu~sex, data=knhanes, alternative=c("two.sided"))  
177  
178
```

Console:

```
      F test to compare two variances  
  
data:  HE_glu by sex  
F = 1.2878, num df = 2978, denom df = 3565, p-value = 5.12e-13  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 1.202255 1.379672  
sample estimates:  
ratio of variances  
      1.287773  
  
> |
```

설명:

귀무가설은 성별과 공복혈당의 모분산이 서로 같다는 것이다. 등분산 검정을 시행하기 위해 script창에 위와 같은 식을 작성했으며, 그 결과가 console 창에 나타났다. $\alpha = 0.05$ 라고 하고 등분산 검정을 시행한 결과 $p\text{-value} < 0.05$ 라는 결론이 도출되었다. 즉, $p\text{-value}$ 가 α 보다 작으므로 귀무가설은 기각된다. 결론적으로 성별과 공복혈당의 모분산은 서로 같지 않다.

3) 등분산을 가정할 때와 가정하지 않을 때에 대하여 각각 t검정을 실시하고 그 결과를 해석하시오.

(귀무가설: 두 모평균이 같다, $\alpha=0.05$)

Script:

```
140
141 #4-3)
142
143 #등분산 가정 t-test
144 t.test(HE_glu~sex, data=knhanes, var.equal=TRUE)
145
146
147 #이분산 가정 t-test
148 t.test(HE_glu~sex, data=knhanes, var.equal=FALSE)
149
150
151
```

Console:

```
Two Sample t-test

data: HE_glu by sex
t = 8.0673, df = 6543, p-value = 8.482e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.621224 5.946038
sample estimates:
mean in group 1 mean in group 2
   103.01947      98.23584

> #이분산 가정 t-test
> t.test(HE_glu~sex, data=knhanes, var.equal=FALSE)

Welch Two Sample t-test

data: HE_glu by sex
t = 7.9768, df = 5989.4, p-value = 1.786e-15
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.608014 5.959249
sample estimates:
mean in group 1 mean in group 2
   103.01947      98.23584

> |
```

설명:

등분산을 가정할 때와 이분산을 가정할 때의 t검정을 각각 시행하였다. 귀무가설은 "두 모평균이 같다"이기 때문에 양측 검정을 시행하였고, alternative를 따로 정하지 않았다. 우선 성별과 공복혈당의 모분산이 같다고 가정하였을 때 $t=8.0673$, $df=6543$, $p\text{-value}=8.482e-16$ 이라는 결과가 도출되었다. 즉, p-value가 0.05보다 작기 때문에 두 모평균이 다르다는 결과가 나왔다. 두 모분산이 다르다고 가정했을 때에는 $t=7.9768$, $df=5989$, $p\text{-value}=1.786e-15$ 이라는 결과가 도출되었다. 등분산을 가정했을 때보다 p-value값이 조금 증가하고 95% 채택역이 넓어지긴 했지만, 여전히 0.05보다 작기에 두 모평균이 같이 않다는 결과가 나온다. 즉, 등분산을 가정했을 때와 가정하지 않았을 때 모두 성별과 공복혈당의 모평균이 같지 않다고 해석할 수 있다.

5. 당뇨병 여부(DE1_pr)가 고혈압 여부(DI1_pr)와 연관성이 유의한지 카이제곱 검정하시오. ($\alpha=0.05$)

1) 당뇨병(DE1_pr)과 고혈압(DI1_pr)의 교차표를 작성하시오.

Script:

```
145
146 #5번
147 ##5-1) xtabs를 사용하여 교차표 작성
148
149
150 ##결측값 처리 이후 교차표
151 contin_5_n=xtabs(data=knhanes, ~DE1_pr_n+DI1_pr_n)
152 contin_5_n
153 ###margin 추가
154 addmargins(contin_5_n)
155 ###빈도 분을 변환
156 prop.table(contin_5_n)
157 ###행별 분을
158 addmargins(prop.table(contin_5_n, margin=1), margin=2)
159 ###열별 분을
160 addmargins(prop.table(contin_5_n, margin=2), margin=1)
161
162
163
```

Console:

```
> ##결측값 처리 이후 교차표
> contin_5_n=xtabs(data=knhanes, ~DE1_pr_n+DI1_pr_n)
> contin_5_n
      DI1_pr_n
DE1_pr_n  0  1
      0   3  9
      1  11 362
> ###margin 추가
> addmargins(contin_5_n)
      DI1_pr_n
DE1_pr_n  0  1 Sum
      0   3  9 12
      1  11 362 373
      Sum 14 371 385
```

```

> ###빈도 분율 변환
> prop.table(contn_5_n)
      DI1_pr_n
DE1_pr_n      0      1
      0 0.007792208 0.023376623
      1 0.028571429 0.940259740
> ###행렬 분율
> addmargins(prop.table(contn_5_n, margin=1), margin=2)
      DI1_pr_n
DE1_pr_n      0      1      Sum
      0 0.250000000 0.750000000 1.000000000
      1 0.02949062 0.97050938 1.000000000
> ###열별 분율
> addmargins(prop.table(contn_5_n, margin=2), margin=1)
      DI1_pr_n
DE1_pr_n      0      1
      0 0.21428571 0.02425876
      1 0.78571429 0.97574124
      Sum 1.00000000 1.00000000
> |

```

설명:

교차표 작성을 위해 'xtabs'을 이용하였으며, 결측값 처리 이후의 교차표를 작성하였다. 결측값 처리 이후 교차표는 "contn_5_n"으로 정의하였으며, margin을 추가한 교차표, 빈도를 분율로 변환한 표, 빈도를 행렬 분율로 변환한 자료와 열별 분율로 변환한 자료를 모두 나타내었다. 그 결과가 console 창에 나타났으며, 표로 정리하면 아래와 같다.

1. 결측값 처리 이후 교차표			
	당뇨병 여부 (DI1_pr)		
고혈압 여부 (DE1_pr)		0	1
	0	3	9
	1	11	362

2. margin 추가				
		당뇨병 여부 (DI1_pr)		
고혈압 (DE1_pr)	여부		0	1
				sum
	0		3	9
	1		11	362
		sum	14	371
				385

3. 빈도 분을 변환			
		당뇨병 여부 (DI1_pr)	
고혈압 여부 (DE1_pr)		0	1
	0	0.007792208	0.023376623
	1	0.028571429	0.940259740

4. 행별 분을 변환				
		당뇨병 여부 (DI1_pr)		
고혈압 (DE1_pr)	여부		0	1
				sum
	0		0.25	0.75
		1	0.02949062	0.97050938
				1

5. 열별 분을 변환			
		당뇨병 여부 (DI1_pr)	
고혈압 여부 (DE1_pr)		0	1
	0	0.21428571	0.02425876
	1	0.78571429	0.97574124
	sum	1	1

2) 카이제곱 검정을 실시하시오.($\alpha=0.05$)

Script:

```
165
166 #5-2) 카이제곱검정 실시
167
168 ##결측값 처리 이후
169 X<-chisq.test(contin_5_n)
170 X
171
172
```

Console:

```
> #5-2) 카이제곱검정 실시
> ##결측값 처리 이후
> X<-chisq.test(contin_5_n)
Warning message:
In chisq.test(contin_5_n) :
  카이제곱 approximation은 정확하지 않을수도 있습니다
> X

      Pearson's Chi-squared test with
      Yates' continuity correction

data:  contin_5_n
X-squared = 10.453, df = 1, p-value =
0.001224

> |
```

설명:

결측값 처리 이후의 카이제곱 검정을 시행하였다. 이때의 귀무가설은 당뇨병 여부와 고혈압 여부가 서로 독립이라는 것이다. 카이제곱 검정 실시 결과 p-value 값이 0.05보다 작았다. 따라서 두 범주가 독립적이라는 귀무가설은 기각되고, 고혈압 여부와 당뇨병 여부가 서로 독립이 아니라는 결론이 도출된다.

3) 각 셀의 기댓값을 구하고 관찰된 값과 유사한지 비교하시오.

Script:

```
240
241 #결측값 처리 이후 관찰빈도
242 contin_5_n
243 #결측값 처리 이후 기대빈도
244 X$expected
245 |
```

Console:

```
> #결측값 처리 이후 관찰빈도
> contin_5_n
      DI1_pr_n
DE1_pr_n  0   1
      0   3   9
      1  11 362
> #결측값 처리 이후 기대빈도
> X$expected
      DI1_pr_n
DE1_pr_n      0      1
      0 0.4363636 11.56364
      1 13.5636364 359.43636
> |
```

설명:

결측값 처리 이후의 관찰 빈도와 기대빈도를 각각 나타내었다. 그 결과 아래와 같은 자료가 console창에 나타났다. 귀무가설 (고혈압 여부와 당뇨병 여부는 서로 확률적 독립이다)이 참일 경우의 기대 빈도는 관찰 빈도와 차이가 있다. 즉, 귀무가설은 기각되며 이는 앞에서 카이제곱 검정을 통해 도출된 결론과 같다.

1. 결측값 처리 이후 관찰 빈도			
	당뇨병 여부 (DI1_pr)		
고혈압 여부 (DE1_pr)		0	1
	0	3	9
	1	11	362

2. 결측값 처리 이후 기대 빈도			
	당뇨병 여부 (DI1_pr)		
고혈압 여부 (DE1_pr)		0	1
	0	0.4363636	11.56364
	1	13.5636364	359.43636

- 4) Fisher's Exact 검정이 필요한지 여부에 대하여 근거를 들어 논하시오. 필요하다고 판정된 경우, Fisher's Exact 검정을 실시하시오($\alpha=0.05$).

Script:

```

181 #5-4)
182 ##결측값 처리 이후 exact test
183 fisher.test(contin_5_n)
184 |

```

Console:

```

> #5-4)
> ##결측값 처리 이후 exact test
> fisher.test(contin_5_n)

      Fisher's Exact Test for Count Data

data:  contin_5_n
p-value = 0.006969
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.655196 52.020680
sample estimates:
odds ratio
 10.78338

> |

```

설명:

Fisher's Exact 검정은 기대빈도가 5보다 작은 셀이 20% 이상을 차지할 때 사용된다. 결측값 처리 후의 기대빈도표를 보면 전체 4개의 셀 중 1개의 셀이 5미만의 값을 가진다. 따라서 이는 20%를 초과하는 25%이기 때문에 Fisher's Exact test가 필요하다.

Fisher's Exact 검정을 시행한 결과 두 변수가 서로 확률적으로 독립이라는 귀무가설이 귀각된다. P-value 값이 0.006969로 0.05보다 작기 때문이다. 따라서 결론적으로 고혈압 여부와 당뇨병 여부는 확률적으로 독립적이지 않다.

6. 위의 내용을 분석한 본인의 R 프로그램을 아래에 텍스트 형태로 제출하세요.(캡처도 무방함)

Script:

```
7
8
9 #1번
10
11 ##import dataset 통해 knhanes.csv 불러오기
12
13 ##ifelse 문을 사용하여 결측값 처리
14 ###고혈압 현재 유병 여부: 8,9 결측값 처리
15 knhanes$DI1_pr_n=as.factor(ifelse(knhanes$DI1_pr<8, knhanes$DI1_pr, NA))
16 ###현재흡연 여부: 9 결측값 처리
17 knhanes$BS3_1_n=as.factor(ifelse(knhanes$BS3_1<9, knhanes$BS3_1, NA))
18 ###평생음주경험: 8,9 결측값 처리
19 knhanes$BD1_n=as.factor(ifelse(knhanes$BD1<8, knhanes$BD1, NA))
20 ###결혼 여부: 9 결측값 처리
21 knhanes$marri_1_n=as.factor(ifelse(knhanes$marri_1<=2, knhanes$marri_1, NA))
22 ###주관적 건강인지: 9 결측값 처리
23 knhanes$D_1_1_n=as.factor(ifelse(knhanes$D_1_1<9, knhanes$D_1_1, NA))
24 ###이상지질혈증 현재 유병 여부: 8,9 결측값 처리
25 knhanes$DI2_pr_n=as.factor(ifelse(knhanes$DI2_pr<8, knhanes$DI2_pr, NA))
26 ###당뇨병 현재 유병 여부: 8,9 결측값 처리
27 knhanes$DE1_pr_n=as.factor(ifelse(knhanes$DE1_pr<8, knhanes$DE1_pr, NA))
28
29
30
31
32
33
34 #2번
35
36 ##psych 패키지 설치 후 사용 준비
37 install.packages("psych")
38 library(psych)
39
40 ##describe로 기술통계 요청
41 ###최종 수축기 혈압
42 describe(knhanes$HE_sbp)
43 ###만나이
44 describe(knhanes$age)
```



```

45 ###신장
46 describe(knhanes$HE_ht)
47 ###체중
48 describe(knhanes$HE_wt)
49 ###허리둘레
50 describe(knhanes$HE_wc)
51 ###공복혈당
52 describe(knhanes$HE_glu)
53 ###총콜레스테롤
54 describe(knhanes$HE_chol)
55 ###EQ-5D index
56 describe(knhanes$EQ5D)
57
58
59
60
61
62 #3번
63
64 ##prettyR 패키지 설치, 사용 준비
65 install.packages("prettyR")
66 library(prettyR)
67
68 ##freq 사용하여 범주형변수 빈도, 퍼센트 요청
69
70 ###고혈압 현재 유병 여부 (결측값 처리 이전)
71 freq(knhanes$DI1_pr)
72 ###고혈압 현재 유병 여부 (결측값 처리 이후)
73 freq(knhanes$DI1_pr_n)
74 ###성별
75 freq(knhanes$sex)
76 ###현재흡연 여부 (결측값 처리 이전)
77 freq(knhanes$BS3_1)
78 ###현재흡연 여부 (결측값 처리 이후)
79 freq(knhanes$BS3_1_n)
80 ###평생음주경험 (결측값 처리 이전)
81 freq(knhanes$BD1)
82 ###평생음주경험 (결측값 처리 이후)
83 freq(knhanes$BD1_n)
84 ###결혼여부 (결측값 처리 이전)
85 freq(knhanes$marri_1)
86 ###결혼여부 (결측값 처리 이후)

```

```

88 ###소득 5분위수 (가구)
89 freq(knhanes$ho_incm5)
90 ###교육수준 재분류 코드
91 freq(knhanes$edu)
92 ###주관적 건강 인지 (결측값 처리 이전)
93 freq(knhanes$D_1_1)
94 ###주관적 건강 인지 (결측값 처리 이후)
95 freq(knhanes$D_1_1_n)
96 ###이상지질혈증 현재 유병 여부 (결측값 처리 이전)
97 freq(knhanes$DI2_pr)
98 ###이상지질혈증 현재 유병 여부 (결측값 처리 이후)
99 freq(knhanes$DI2_pr_n)
100 ###당뇨병 현재 유병 여부 (결측값 처리 이전)
101 freq(knhanes$DE1_pr)
102 ###당뇨병 현재 유병 여부 (결측값 처리 이후)
103 freq(knhanes$DE1_pr_n)
104
105
106
107 |
108
109
110
111 #4번
112
113 #4-1)
114
115 #psych 패키지 다시 실행
116 library(psych)
117
118 #describeBy를 통해 sex에 따른 HE_glu 구하기
119 describeBy(knhanes$HE_glu, knhanes$sex)
120
121 #boxplot 이용하여 상자그림 그리기
122 boxplot(HE_glu~sex, data=knhanes, col="green")
123

```

```

125 #4-2)
126
127 #성별에 따른 공복혈당 등분산 검정
128 var.test(HE_glu~sex, data=knhanes, alternative=c("two.sided"))
129
130
131
132
133 #4-3)
134
135 #등분산 가정 t-test
136 t.test(HE_glu~sex, data=knhanes, var.equal=TRUE)
137
138
139 #이분산 가정 t-test
140 t.test(HE_glu~sex, data=knhanes, var.equal=FALSE)
141
142

```

```

145
146 #5번
147 ##5-1) xtabs를 사용하여 교차표 작성
148
149
150 ##결측값 처리 이후 교차표
151 contin_5_n=xtabs(data=knhanes, ~DE1_pr_n+DI1_pr_n)
152 contin_5_n
153 ###margin 추가
154 addmargins(contin_5_n)
155 ###빈도 분을 변환
156 prop.table(contin_5_n)
157 ###행별 분을
158 addmargins(prop.table(contin_5_n, margin=1), margin=2)
159 ###열별 분을
160 addmargins(prop.table(contin_5_n, margin=2), margin=1)
161
162
163
164
165
166 #5-2) 카이제곱검정 실시
167 ##결측값 처리 이후
168 X<-chisq.test(contin_5_n)
169 X
170
171

```

```

172
173 #5-3)
174
175 ##결측값 처리 이후 관찰빈도
176 contin_5_n
177 ##결측값 처리 이후 기대빈도
178 X$expected
179
180
181 #5-4)
182 ##결측값 처리 이후 exact test
183 fisher.test(contin_5_n)
184
185

```

Console:

1.

```
> #1번
> ##import dataset 통해 knhanes.csv 불러오기
> ##ifelse 문을 사용하여 결측값 처리
> ###고혈압 현재 유병 여부: 8,9 결측값 처리
> knhanes$DI1_pr_n=as.factor(ifelse(knhanes$DI1_pr<8, knhanes$DI1_pr, NA))
> ###현재흡연 여부: 9 결측값 처리
> knhanes$BS3_1_n=as.factor(ifelse(knhanes$BS3_1<9, knhanes$BS3_1, NA))
> ###평생음주경험: 8,9 결측값 처리
> knhanes$BD1_n=as.factor(ifelse(knhanes$BD1<8, knhanes$BD1, NA))
> ###결혼 여부: 9 결측값 처리
> knhanes$marri_1_n=as.factor(ifelse(knhanes$marri_1<=2, knhanes$marri_1, NA))
> ###주관적 건강인지: 9 결측값 처리
> knhanes$D_1_1_n=as.factor(ifelse(knhanes$D_1_1<9, knhanes$D_1_1, NA))
> ###이상지질혈증 현재 유병 여부: 8,9 결측값 처리
> knhanes$DI2_pr_n=as.factor(ifelse(knhanes$DI2_pr<8, knhanes$DI2_pr, NA))
> ###당뇨병 현재 유병 여부: 8,9 결측값 처리
> knhanes$DE1_pr_n=as.factor(ifelse(knhanes$DE1_pr<8, knhanes$DE1_pr, NA))
```

2.

```
> ##psych 패키지 설치 후 사용 준비
> install.packages("psych")
```

```
> library(psych)
> ##describe로 기술통계 요청
> ###최종 수축기 혈압
> describe(knhanes$HE_sbp)
  vars    n  mean    sd median trimmed
X1    1 6878 117.97 16.68   116  116.73
      mad min max range skew kurtosis  se
X1 16.31  77 232   155 0.94    1.95 0.2
> ###만나이
> describe(knhanes$age)
  vars    n  mean    sd median trimmed
X1    1 8127 43.12 22.73    45  43.58
      mad min max range skew kurtosis  se
X1 26.69   1  80   79 -0.2   -1.06 0.25
> ###신장
> describe(knhanes$HE_ht)
  vars    n  mean    sd median trimmed
X1    1 7684 157.42 18.65  160.8  160.48
      mad min max range skew kurtosis
X1 11.42  73.9 194 120.1 -1.91    4.34
      se
X1 0.21
> ###체중
> describe(knhanes$HE_wt)
  vars    n  mean    sd median trimmed
X1    1 7695 58.37 17.79   59.5  59.55
      mad min max range skew kurtosis  se
X1 13.49   7.9 140 132.1 -0.49    0.92 0.2
> ###허리둘레
> describe(knhanes$HE_wc)
  vars    n  mean    sd median trimmed
X1    1 7662 77.7 13.79   79.05   78.4
      mad min max range skew kurtosis
X1 13.12  24.5 128.3 103.8 -0.37    0.09
      se
X1 0.16
> ###공복혈당
> describe(knhanes$HE_glu)
  vars    n  mean    sd median trimmed
X1    1 6545 100.41 24.01    95  96.23
      mad min max range skew kurtosis  se
X1 10.38   54 379   325 4.39   29.63 0.3
```

```

> ###총콜레스테롤
> describe(knhanes$HE_chol)
  vars    n mean    sd median trimmed
X1    1 6546 190.4 38.09    188 189.04
      mad min max range skew kurtosis  se
X1 37.06  64 451   387 0.42    0.56 0.47
> ###EQ-5D index
> describe(knhanes$EQ5D)
  vars    n mean    sd median trimmed mad
X1    1 5833 0.95 0.11     1    0.97  0
      min max range  skew kurtosis se
X1 -0.17   1  1.17 -3.47   17.99  0

```

3.

```

> library(prettyR)
> ##freq 사용하여 범주형변수 빈도, 퍼센트 요청
> ###고혈압 현재 유병 여부 (결측값 처리 이전)
> freq(knhanes$DI1_pr)

Frequencies for knhanes$DI1_pr
      8    1    0    9  NA
6169 1449   79    2 428
%    75.9 17.8    1    0 5.3
%!NA 80.1 18.8    1    0

> ###고혈압 현재 유병 여부 (결측값 처리 이후)
> freq(knhanes$DI1_pr_n)

Frequencies for knhanes$DI1_pr_n
      1    0  NA
1449   79 6599
%    17.8    1 81.2
%!NA 94.8  5.2

> ###성별
> freq(knhanes$sex)

Frequencies for knhanes$sex
      2    1  NA
4361 3766    0
%    53.7 46.3    0
%!NA 53.7 46.3

```

```

> ###현재흡연 여부 (결측값 처리 이전)
> freq(knhanes$BS3_1)

Frequencies for knhanes$BS3_1
      8    3    1    2    9  NA
5199 1331  941  153   75 428
%    64 16.4 11.6  1.9 0.9 5.3
%!NA 67.5 17.3 12.2    2    1

> ###현재흡연 여부 (결측값 처리 이후)
> freq(knhanes$BS3_1_n)

Frequencies for knhanes$BS3_1_n
      8    3    1    2  NA
5199 1331  941  153 503
%    64 16.4 11.6  1.9 6.2
%!NA 68.2 17.5 12.3    2

```

```
> ###평생음주경험 (결측값 처리 이전)
```

```
> freq(knhanes$BD1)
```

```
Frequencies for knhanes$BD1
```

	2	1	8	9	NA
	5561	1083	990	65	428
%	68.4	13.3	12.2	0.8	5.3
%!NA	72.2	14.1	12.9	0.8	

```
> ###평생음주경험 (결측값 처리 이후)
```

```
> freq(knhanes$BD1_n)
```

```
Frequencies for knhanes$BD1_n
```

	2	1	NA
	5561	1083	1483
%	68.4	13.3	18.2
%!NA	83.7	16.3	

```
> ###결혼여부 (결측값 처리 이전)
```

```
> freq(knhanes$marri_1)
```

```
Frequencies for knhanes$marri_1
```

	1	2	NA
	5386	2740	1
%	66.3	33.7	0
%!NA	66.3	33.7	

```
> ###결혼여부 (결측값 처리 이후)
```

```
> freq(knhanes$marri_1_n)
```

```
Frequencies for knhanes$marri_1_n
```

	1	2	NA
	5386	2740	1
%	66.3	33.7	0
%!NA	66.3	33.7	

```
> ###소득 5분위수 (가구)
```

```
> freq(knhanes$ho_incm5)
```

```
Frequencies for knhanes$ho_incm5
```

	5	3	4	2	1	NA
	1900	1842	1830	1431	1096	28
%	23.4	22.7	22.5	17.6	13.5	0.3
%!NA	23.5	22.7	22.6	17.7	13.5	

```
> ###교육수준 재분류 코드
```

```
> freq(knhanes$edu)
```

```
Frequencies for knhanes$edu
```

	1	4	3	2	NA
	2501	2212	1812	804	798
%	30.8	27.2	22.3	9.9	9.8
%!NA	34.1	30.2	24.7	11	

```
> ###주관적 건강 인지 (결측값 처리 이전)
```

```
> freq(knhanes$D_1_1)
```

```
Frequencies for knhanes$D_1_1
```

	3	2	4	1	9	5	NA
	3565	1996	975	590	334	239	428
%	43.9	24.6	12	7.3	4.1	2.9	5.3
%!NA	46.3	25.9	12.7	7.7	4.3	3.1	


```
> ###주관적 건강 인지 (결측값 처리 이후)
> freq(knhanes$D_1_1_n)
```

```
Frequencies for knhanes$D_1_1_n
      3      2      4      1      5      NA
3565 1996  975  590  239  762
%   43.9 24.6  12  7.3  2.9  9.4
%!NA 48.4 27.1 13.2   8  3.2
```

```
> ###이상지질혈증 현재 유병 여부 (결측값 처리 이전)
> freq(knhanes$DI2_pr)
```

```
Frequencies for knhanes$DI2_pr
      8      1      0      9      NA
6530  940  227      2  428
%   80.3 11.6  2.8   0  5.3
%!NA 84.8 12.2  2.9   0
```

```
> ###이상지질혈증 현재 유병 여부 (결측값 처리 이후)
> freq(knhanes$DI2_pr_n)
```

```
Frequencies for knhanes$DI2_pr_n
      1      0      NA
 940  227 6960
%   11.6  2.8 85.6
%!NA 80.5 19.5
```

```
> ###당뇨병 현재 유병 여부 (결측값 처리 이전)
> freq(knhanes$DE1_pr)
```

```
Frequencies for knhanes$DE1_pr
      8      1      0      9      NA
7100  579   16      4  428
%   87.4  7.1  0.2   0  5.3
%!NA 92.2  7.5  0.2  0.1
```

```
> ###당뇨병 현재 유병 여부 (결측값 처리 이후)
> freq(knhanes$DE1_pr_n)
```

```
Frequencies for knhanes$DE1_pr_n
      1      0      NA
 579   16 7532
%    7.1  0.2 92.7
%!NA 97.3  2.7
```

4.

```
> #4번
> #4-1)
> #psych 패키지 다시 실행
> library(psych)
> #describeBy를 통해 sex에 따른 HE_glu 구하기
> describeBy(knhanes$HE_glu, knhanes$sex)

Descriptive statistics by group
group: 1
  vars   n  mean   sd median trimmed
X1     1 2979 103.02 25.49    97   98.55
      mad min max range skew kurtosis  se
X1 11.86  54 352   298  3.95   22.72 0.47
-----
group: 2
  vars   n  mean   sd median trimmed
X1     1 3566  98.24 22.46    93   94.36
      mad min max range skew kurtosis  se
X1  8.9  56 379   323  4.91   39.01 0.38
> #boxplot 이용하여 상자그림 그리기
> boxplot(HE_glu~sex, data=knhanes, col="green")
> #4-2)
> #성별에 따른 공복혈당 등분산 검정
> var.test(HE_glu~sex, data=knhanes, alternative=c("two.sided"))

      F test to compare two variances

data:  HE_glu by sex
F = 1.2878, num df = 2978, denom df =
3565, p-value = 5.12e-13
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.202255 1.379672
sample estimates:
ratio of variances
 1.287773
```

```
> #4-3)
> #등분산 가정 t-test
> t.test(HE_glu~sex, data=knhanes, var.equal=TRUE)

      Two Sample t-test

data:  HE_glu by sex
t = 8.0673, df = 6543, p-value =
8.482e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.621224 5.946038
sample estimates:
mean in group 1 mean in group 2
 103.01947      98.23584

> #이분산 가정 t-test
> t.test(HE_glu~sex, data=knhanes, var.equal=FALSE)

      Welch Two Sample t-test

data:  HE_glu by sex
t = 7.9768, df = 5989.4, p-value =
1.786e-15
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.608014 5.959249
sample estimates:
mean in group 1 mean in group 2
 103.01947      98.23584
```

5.

```
> #5번
> ##5-1) xtabs를 사용하여 교차표 작성
> ##결측값 처리 이후 교차표
> contin_5_n=xtabs(data=knhanes, ~DE1_pr_n+DI1_pr_n)
> contin_5_n
      DI1_pr_n
DE1_pr_n  0   1
      0   3   9
      1  11 362
> ###margin 추가
> addmargins(contin_5_n)
      DI1_pr_n
DE1_pr_n  0   1 Sum
      0   3   9 12
      1  11 362 373
      Sum 14 371 385
> ###빈도 분을 변환
> prop.table(contin_5_n)
      DI1_pr_n
DE1_pr_n      0      1
      0 0.007792208 0.023376623
      1 0.028571429 0.940259740
> ###행별 분을
> addmargins(prop.table(contin_5_n, margin=1), margin=2)
      DI1_pr_n
DE1_pr_n      0      1      Sum
      0 0.25000000 0.75000000 1.00000000
      1 0.02949062 0.97050938 1.00000000
> ###열별 분을
> addmargins(prop.table(contin_5_n, margin=2), margin=1)
      DI1_pr_n
DE1_pr_n      0      1
      0 0.21428571 0.02425876
      1 0.78571429 0.97574124
      Sum 1.00000000 1.00000000
```

```

> #5-2) 카이제곱검정 실시
> ##결측값 처리 이후
> X<-chisq.test(contin_5_n)
Warning message:
In chisq.test(contin_5_n) :
  카이제곱 approximation은 정확하지 않을수도 있습니다
> X

      Pearson's Chi-squared test with
      Yates' continuity correction

data:  contin_5_n
X-squared = 10.453, df = 1, p-value =
0.001224

> #5-3)
> ##결측값 처리 이후 관찰빈도
> contin_5_n
      DI1_pr_n
DE1_pr_n  0   1
      0   3   9
      1  11 362
> ##결측값 처리 이후 기대빈도
> X$expected
      DI1_pr_n
DE1_pr_n      0      1
      0  0.4363636 11.56364
      1 13.5636364 359.43636

> #5-4)
> ##결측값 처리 이후 exact test
> fisher.test(contin_5_n)

      Fisher's Exact Test for Count Data

data:  contin_5_n
p-value = 0.006969
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.655196 52.020680
sample estimates:
odds ratio
 10.78338

```