

CH2

2023-06-06

5. Factor 다루기

5.1 Factor로 변환하기

데이터에서 자료의 형태는 연속형도 있지만, 범주형 변수도 있음
R에서는 범주형을 factor형태로 다루게 됨

```
library(tidyverse)
dat <- read.csv("C:\\Users\\ph102\\Desktop\\P\\bio\\ch2\\Ch2_chb.csv")
```

```
dat1<-dat
class(dat1$gender)
```

```
## [1] "character"
```

```
table(dat1$gender)
```

```
##
##  F  M
##  6 18
```

```
dat1$sex1<-as.factor(dat1$gender)
class(dat1$sex1)
```

```
## [1] "factor"
```

```
table(dat1$sex1)
```

```
##
##  F  M
##  6 18
```

```
dat1$sex2<-as_factor(dat1$gender)
class(dat1$sex2)
```

```
## [1] "factor"
```

```
table(dat1$sex2)
```

```
##
##  M  F
## 18  6
```

as.factor, factor와 as_factor의 차이는 F, M의 순서
여기서 level의 개념이 적용됨

5.2 Factor level 이해하기

factor에서 level이란 범주형 변수값의 순서

```
dat1$risk_gr1<-as.factor(dat1$risk_gr)
levels(dat1$risk_gr1)
```

```
## [1] "high"          "intermediate" "low"
```

```
dat1$risk_gr2<-factor(dat1$risk_gr,
                      levels=c('low','intermediate','high'))
levels(dat1$risk_gr2)
```

```
## [1] "low"           "intermediate" "high"
```

5.3 forcat 패키지

```
dat1$region<-factor(dat1$region)
table(dat1$region)
```

```
##
##   busan   daegu   daejun gwangju incheon   jeju   sejong   seoul
##      4      4      2      2      2      1      1      8
```

```
dat1$region1<-fct_infreq(dat1$region)
levels(dat1$region1)
```

```
## [1] "seoul"   "busan"   "daegu"   "daejun"  "gwangju" "incheon" "jeju"
## [8] "sejong"
```

```
dat1$region2 <-fct_recode(dat1$region,
                          'north' = 'seoul',
                          'north' = 'incheon',
                          'south' = 'gwangju',
                          'south' = 'jeju',
                          'south' = 'sejong',
                          'south' = 'daejun',
                          'east' = 'daegu',
                          'east' = 'busan')
```

```
dat1$region3 <- fct_collapse(dat1$region,
                              'north' = c('seoul','incheon'),
                              'south' = c('gwangju','jeju','sejong','daejun'),
                              'east' = c('daegu','busan'))

dat1 %>%
  count(region, region2, region3)
```

```
##   region region2 region3 n
## 1  busan    east    east  4
## 2  daegu    east    east  4
## 3  daejun   south   south  2
## 4 gwangju   south   south  2
## 5 incheon   north   north  2
## 6  jeju     south   south  1
## 7 sejong    south   south  1
## 8 seoul     north   north  8
```

```
dat1$region4 <- fct_lump(dat1$region, n=4)
dat1 %>%
  count(region,region4)
```

```
##   region region4 n
## 1  busan   busan  4
## 2  daegu   daegu  4
## 3  daejun  daejun  2
## 4 gwangju gwangju  2
## 5 incheon incheon  2
## 6  jeju    Other   1
## 7 sejong   Other   1
## 8  seoul   seoul   8
```

6. 기술통계

```
Hmisc::describe(dat$age)
```

```
## dat$age
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      24      0       15    0.987    45.54    9.576    29.15    30.60
##      .25     .50     .75     .90     .95
##    42.50    49.00    50.25    51.70    53.70
##
## Value      26.00 28.88 29.96 31.76 32.84 40.76 42.92 44.72 47.96 48.68 49.76
## Frequency      1      1      1      1      1      1      1      1      2      3      5
## Proportion 0.042 0.042 0.042 0.042 0.042 0.042 0.042 0.042 0.042 0.083 0.125 0.208
##
## Value      50.84 51.92 53.72 62.00
## Frequency      3      1      1      1
## Proportion 0.125 0.042 0.042 0.042
##
## For the frequency table, variable is rounded to the nearest 0.36
```

7. 데이터 수정 및 결측치

```
dat[dat$dna=='undetectable',]
```

```
##      id index_date gender age  last_date treat_gr lc hcc  hcc_date  hcc_yr
## 19 19 2007/02/15      M 48 2017/08/25      TDF  0  0 2017/08/25 10.67778
##      lab_date alt bil inr cr plt alb eag      dna dna_log risk_gr region
## 19 2007/02/01 34 1.1 1.1 1 NA 4.3 0 undetectable      1      low sejong
```

```
dat[19, 'dna'] <- '0'
dat$dna
```

```
## [1] "31934937" "1100" "35968299" "9145608" "1095481"
## [6] "900" "29000000" "1400000000" "2000000" "58425664"
## [11] "12340762" "47000" "220" "14000000" "497946000"
## [16] "720" "82000000" "440000" "0" "24000"
## [21] "66393" "337540995" "280000" "270000"
```

```
class(dat$dna)
```

```
## [1] "character"
```

```
dat$dna <- as.numeric(dat$dna)
class(dat$dna)
```

```
## [1] "numeric"
```

7.1 결측값 확인

```
mean(dat$alb)
```

```
## [1] NA
```

```
mean(dat$alb, na.rm=T)
```

```
## [1] 3.7
```

```
is.na(dat$alb)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
sum(is.na(dat$alb))
```

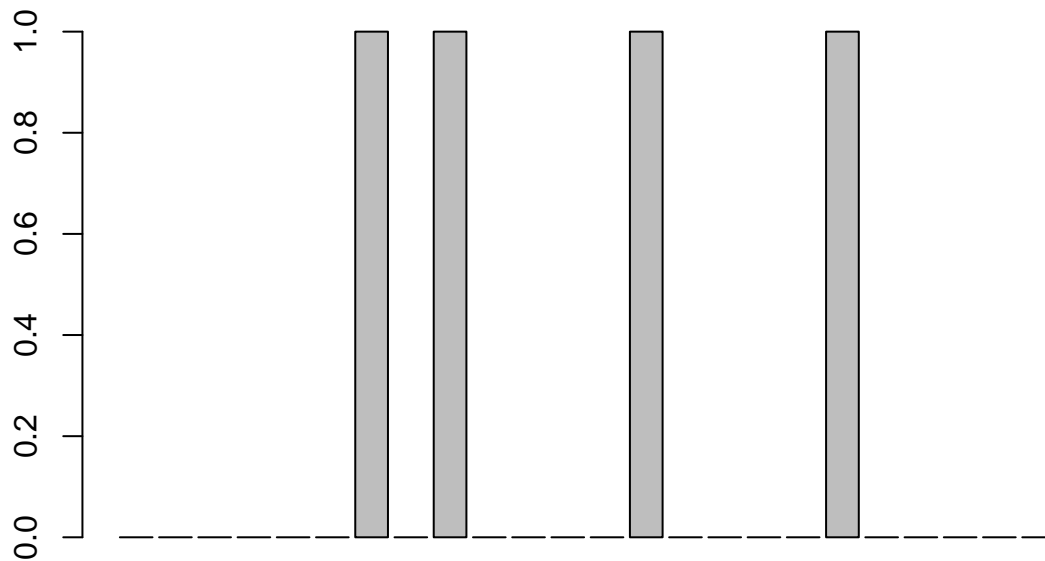
```
## [1] 2
```


7.2 결측값 한 번에 확인

```
rowSums(is.na(dat))
```

```
## [1] 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0
```

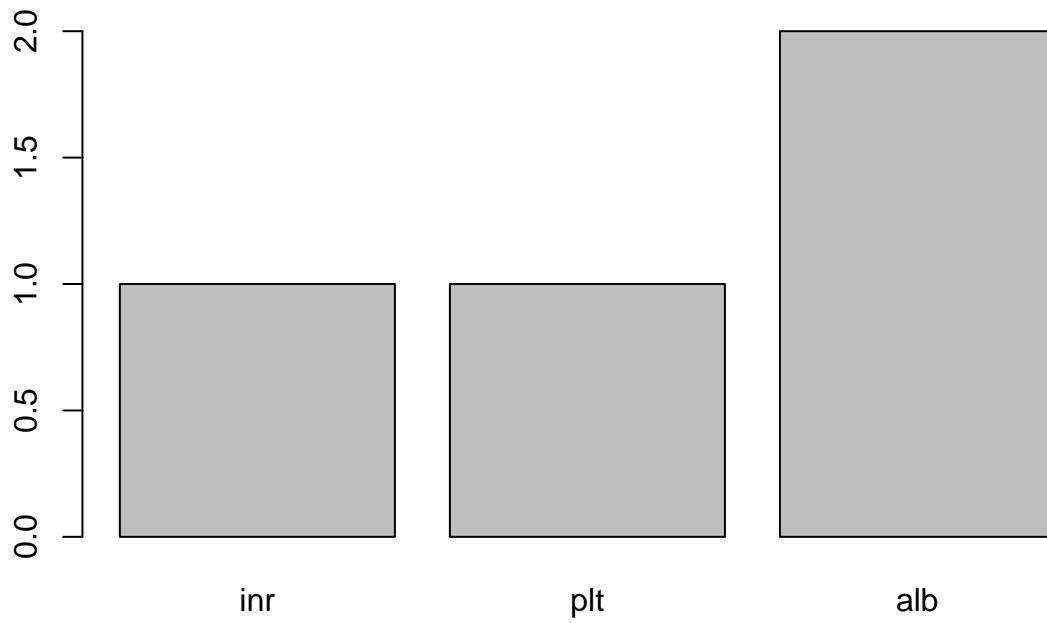
```
barplot(rowSums(is.na(dat)))
```



```
na.count <- apply(dat, 2, function(x) sum(is.na(x)))  
na.count
```

```
##      id index_date  gender    age last_date treat_gr    lc  
##      0         0      0      0         0         0      0  
##      hcc  hcc_date hcc_yr lab_date    alt      bil    inr  
##      0         0      0      0         0         0      1  
##      cr      plt    alb    eag    dna  dna_log risk_gr  
##      0         1      2      0         0         0      0  
##      region  
##      0
```

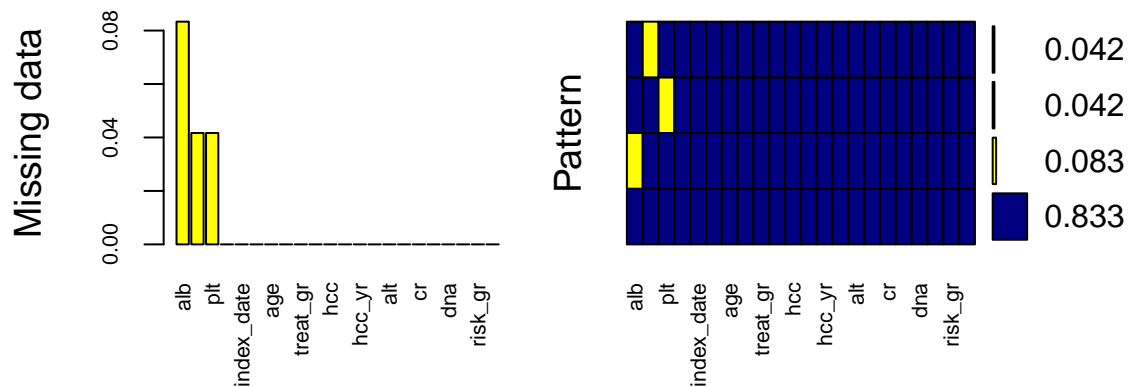
```
barplot(na.count[na.count>0])
```



7.3 패키지 이용

```
library(VIM)

missing <- aggr(dat, col=c('navyblue','yellow'),
               numbers=TRUE, sortVars=TRUE,
               labels=names(dat1), cex.axis=.7,
               gap=3,
               ylab=c('Missing data','Pattern'))
```



```
##
## Variables sorted by number of missings:
## Variable      Count
## alb 0.08333333
## inr 0.04166667
## plt 0.04166667
## id 0.00000000
## index_date 0.00000000
## gender 0.00000000
## age 0.00000000
## last_date 0.00000000
## treat_gr 0.00000000
## lc 0.00000000
## hcc 0.00000000
## hcc_date 0.00000000
## hcc_yr 0.00000000
## lab_date 0.00000000
## alt 0.00000000
## bil 0.00000000
## cr 0.00000000
## eag 0.00000000
## dna 0.00000000
## dna_log 0.00000000
## risk_gr 0.00000000
## region 0.00000000
```

```
missing
```

```
##  
## Missings in variables:  
## Variable Count  
##      inr      1  
##      plt      1  
##      alb      2
```

8. apply 함수

8.1 tapply

```
tapply(dat$age, dat$gender, mean)
```

```
##           F           M  
## 51.16667 43.66667
```

```
dat %>%  
  group_by(gender) %>%  
  summarise(mean(age))
```

```
## # A tibble: 2 x 2  
##   gender `mean(age)`  
##   <chr>      <dbl>  
## 1 F          51.2  
## 2 M          43.7
```

8.2 sapply

```
sapply(dat, class)
```

```
##      id index_date  gender      age last_date  treat_gr  
## "integer" "character" "character" "integer" "character" "character"  
##      lc      hcc  hcc_date  hcc_yr  lab_date      alt  
## "integer" "integer" "character" "numeric" "character" "integer"  
##      bil      inr      cr      plt      alb      eag  
## "numeric" "numeric" "numeric" "integer" "numeric" "integer"  
##      dna      dna_log  risk_gr  region  
## "numeric" "numeric" "character" "character"
```

8.3 lapply

```
lapply(dat[,c('age', 'gender')], class)
```

```
## $age  
## [1] "integer"  
##  
## $gender  
## [1] "character"
```

10. 중복 결과값 다루기

10.1 unique 함수 사용

```
dat$cr
```

```
## [1] 1.00 0.60 1.05 0.93 0.90 1.20 0.91 0.70 0.73 1.20 1.10 0.81 1.00 0.69 1.00  
## [16] 0.92 0.81 0.60 1.00 0.63 0.80 0.99 0.91 0.89
```

```
length(dat$cr)
```

```
## [1] 24
```

```
unique(dat$cr)
```

```
## [1] 1.00 0.60 1.05 0.93 0.90 1.20 0.91 0.70 0.73 1.10 0.81 0.69 0.92 0.63 0.80  
## [16] 0.99 0.89
```

```
length(unique(dat$cr))
```

```
## [1] 17
```

```
unique(dat$gender, dat$cr)
```

```
## [1] "M" "F"
```

```
unique(dat[,c('gender','cr')])
```

```
##      gender    cr
## 1         M 1.00
## 2         F 0.60
## 3         M 1.05
## 4         M 0.93
## 5         M 0.90
## 6         M 1.20
## 7         M 0.91
## 8         F 0.70
## 9         F 0.73
## 11        M 1.10
## 12        M 0.81
## 14        M 0.69
## 16        M 0.92
## 17        F 0.81
## 20        F 0.63
## 21        M 0.80
## 22        M 0.99
## 24        M 0.89
```

```
dat %>%  
  distinct(gender, cr)
```

```
##   gender    cr  
## 1      M 1.00  
## 2      F 0.60  
## 3      M 1.05  
## 4      M 0.93  
## 5      M 0.90  
## 6      M 1.20  
## 7      M 0.91  
## 8      F 0.70  
## 9      F 0.73  
## 10     M 1.10  
## 11     M 0.81  
## 12     M 0.69  
## 13     M 0.92  
## 14     F 0.81  
## 15     F 0.63  
## 16     M 0.80  
## 17     M 0.99  
## 18     M 0.89
```


10.2 Duplicated 함수 사용

```
dat$gender
```

```
## [1] "M" "F" "M" "M" "M" "M" "M" "F" "F" "M" "M" "M" "M" "M" "M" "M" "F" "F" "M"  
## [20] "F" "M" "M" "M" "M"
```

```
duplicated(dat$gender)
```

```
## [1] FALSE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [13] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
temp<-rbind(dat, dat[c(3,6,10),])
```

```
duplicated(temp$id)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [25] TRUE TRUE TRUE
```

```
!duplicated(temp$id)
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [13] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
## [25] FALSE FALSE FALSE
```

```
temp1<-temp[!duplicated(temp$id),]  
unique(temp1$id)
```

```
## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
```

11. Table 1 만들기

11.1 moonBook 패키지

```
library(moonBook)
mytable(treat_gr~age+gender+lc+alt+bil+
        inr+cr+plt+alb+eag+dna_log,
        data=dat)
```

```
##
##   Descriptive Statistics by 'treat_gr'
## -----
##           ETV           TDF           p
##           (N=15)        (N=9)
## -----
## age      43.7 ± 9.2    48.6 ± 8.5    0.214
## gender
##   - F      4 (26.7%)    2 (22.2%)
##   - M     11 (73.3%)    7 (77.8%)
## lc
##   - 0      6 (40.0%)    4 (44.4%)
##   - 1      9 (60.0%)    5 (55.6%)
## alt      94.2 ± 93.1  123.0 ± 141.1  0.552
## bil       1.3 ± 0.4    1.2 ± 0.4    0.686
## inr       1.1 ± 0.2    1.1 ± 0.0    0.416
## cr        0.9 ± 0.2    0.9 ± 0.1    0.769
## plt      158.4 ± 56.9  146.5 ± 36.0    0.599
## alb       3.6 ± 0.6    3.9 ± 0.5    0.252
## eag
##   - 0      5 (33.3%)    4 (44.4%)
##   - 1     10 (66.7%)    5 (55.6%)
## dna_log   5.8 ± 2.2    6.0 ± 2.4    0.811
## -----
```

```
mytable(treat_gr~age+gender+lc+alt+bil+
        inr+cr+plt+alb+eag+dna_log,
        data=dat, method=2)
```

```
##
##      Descriptive Statistics by 'treat_gr'
## -----
##              ETV              TDF              p
##              (N=15)            (N=9)
## -----
## age          49.0 [37.0;50.5]    49.0 [48.0;50.0]    0.548
## gender
##   - F         4 (26.7%)          2 (22.2%)
##   - M        11 (73.3%)          7 (77.8%)
## lc
##   - 0         6 (40.0%)          4 (44.4%)
##   - 1         9 (60.0%)          5 (55.6%)
## alt          61.0 [33.0;95.5]    94.0 [48.0;106.0]  0.387
## bil           1.3 [ 1.0; 1.4]     1.1 [ 1.0; 1.5]    0.589
## inr           1.1 [ 1.0; 1.2]     1.1 [ 1.0; 1.1]    0.899
## cr            0.9 [ 0.7; 1.0]     0.9 [ 0.8; 0.9]    0.654
## plt          164.0 [137.5;183.5]  156.0 [119.0;163.5] 0.366
## alb           3.8 [ 3.2; 4.1]     4.0 [ 3.7; 4.2]    0.181
## eag
##   - 0         5 (33.3%)          4 (44.4%)
##   - 1        10 (66.7%)          5 (55.6%)
## dna_log       6.3 [ 3.7; 7.3]     6.0 [ 4.8; 7.6]    0.726
## -----
```

```
mytable(treat_gr~age+gender+lc+alt+bil+
        inr+cr+plt+alb+eag+dna_log,
        data=dat, show.total=T)
```

```
##
##      Descriptive Statistics by 'treat_gr'
## -----
##           ETV           TDF           Total           p
##           (N=15)        (N=9)        (N=24)
## -----
## age      43.7 ± 9.2    48.6 ± 8.5    45.5 ± 9.1    0.214
## gender
##   - F      4 (26.7%)    2 (22.2%)    6 (25.0%)
##   - M     11 (73.3%)    7 (77.8%)   18 (75.0%)
## lc
##   - 0      6 (40.0%)    4 (44.4%)   10 (41.7%)
##   - 1      9 (60.0%)    5 (55.6%)   14 (58.3%)
## alt     94.2 ± 93.1  123.0 ± 141.1 105.0 ± 111.4 0.552
## bil      1.3 ± 0.4    1.2 ± 0.4    1.3 ± 0.4    0.686
## inr      1.1 ± 0.2    1.1 ± 0.0    1.1 ± 0.2    0.416
## cr       0.9 ± 0.2    0.9 ± 0.1    0.9 ± 0.2    0.769
## plt     158.4 ± 56.9 146.5 ± 36.0 154.3 ± 50.0 0.599
## alb      3.6 ± 0.6    3.9 ± 0.5    3.7 ± 0.6    0.252
## eag
##   - 0      5 (33.3%)    4 (44.4%)    9 (37.5%)
##   - 1     10 (66.7%)    5 (55.6%)   15 (62.5%)
## dna_log  5.8 ± 2.2    6.0 ± 2.4    5.9 ± 2.2    0.811
## -----
```

```
mytable(~age+gender+lc+alt+bil+
        inr+cr+plt+alb+eag+dna_log,
        data=dat)
```

```
##
##          Descriptive Statistics
## -----
##          Mean ± SD or %    N  Missing (%)
## -----
##    age          45.5 ± 9.1  24  0  ( 0.0%)
##    gender
##      - F          6  (25.0%)
##      - M          18  (75.0%)
##    lc              24  0  ( 0.0%)
##      - 0          10  (41.7%)
##      - 1          14  (58.3%)
##    alt        105.0 ± 111.4  24  0  ( 0.0%)
##    bil           1.3 ± 0.4  24  0  ( 0.0%)
##    inr           1.1 ± 0.2  23  1  ( 4.2%)
##    cr            0.9 ± 0.2  24  0  ( 0.0%)
##    plt        154.3 ± 50.0  23  1  ( 4.2%)
##    alb           3.7 ± 0.6  22  2  ( 8.3%)
##    eag              24  0  ( 0.0%)
##      - 0           9  (37.5%)
##      - 1          15  (62.5%)
##    dna_log        5.9 ± 2.2  24  0  ( 0.0%)
## -----
```

```
table1 <-mytable(treat_gr~age+gender+lc+alt+bil+  
  inr+cr+plt+alb+eag+dna_log,  
  data=dat, method=2)  
mycsv(table1, file='C:\\Users\\phl02\\Desktop\\P\\bio\\ch2\\table1.csv',row.names=FALSE)
```

11.2 tableone 패키지

```
library(tableone)
listVars <- names(dat[, c('age','gender','lc','alt','bil',
                          'inr','cr','plt','alb','eag','dna_log')])
catVars <- c('gender','lc')
table1 <- CreateTableOne(vars = listVars,
                          factorVars = catVars,
                          strata = c('treat_gr'),
                          data = dat)

table1
```

		Stratified by treat_gr			
		ETV	TDF	p	test
##	n	15	9		
##	age (mean (SD))	43.73 (9.16)	48.56 (8.52)	0.214	
##	gender = M (%)	11 (73.3)	7 (77.8)	1.000	
##	lc = 1 (%)	9 (60.0)	5 (55.6)	1.000	
##	alt (mean (SD))	94.20 (93.09)	123.00 (141.11)	0.552	
##	bil (mean (SD))	1.29 (0.41)	1.22 (0.42)	0.686	
##	inr (mean (SD))	1.11 (0.19)	1.07 (0.04)	0.504	
##	cr (mean (SD))	0.90 (0.20)	0.88 (0.11)	0.769	
##	plt (mean (SD))	158.40 (56.88)	146.50 (36.00)	0.599	
##	alb (mean (SD))	3.59 (0.64)	3.90 (0.52)	0.252	
##	eag (mean (SD))	0.67 (0.49)	0.56 (0.53)	0.605	
##	dna_log (mean (SD))	5.78 (2.18)	6.01 (2.41)	0.811	

```
summary(table1)
```

```
##
##      ### Summary of continuous variables ###
##
## treat_gr: ETV
##      n miss p.miss  mean    sd median   p25 p75  min max skew kurt
## age      15   0      0  43.7  9.2   49.0  37.0  50 26.0  54 -0.8 -0.8
## alt      15   0      0  94.2 93.1   61.0  33.0  96 31.0 359  2.1  4.2
## bil      15   0      0   1.3  0.4    1.3   1.0   1  0.7   2  0.9  0.5
## inr      15   1      7   1.1  0.2    1.1   1.0   1  0.9   2  1.2  1.3
## cr       15   0      0   0.9  0.2    0.9   0.7   1  0.6   1 -0.2 -1.1
## plt      15   0      0 158.4 56.9  164.0 137.5 184 55.0 292  0.2  1.7
## alb      15   1      7   3.6  0.6    3.8   3.2   4  2.1   4 -1.1  0.7
## eag      15   0      0   0.7  0.5    1.0   0.0   1  0.0   1 -0.8 -1.6
## dna_log  15   0      0   5.8  2.2    6.3   3.7   7  2.3   9 -0.4 -1.3
## -----
## treat_gr: TDF
##      n miss p.miss  mean    sd median   p25 p75  min max skew kurt
## age      9   0      0  48.6 9e+00   49.0  48.0  50.0 29.0  62 -1.30  4.55
## alt      9   0      0 123.0 1e+02   94.0  48.0 106.0 34.0 486  2.61  7.27
## bil      9   0      0   1.2 4e-01    1.1   1.0   1.5  0.7   2  0.63 -0.83
## inr      9   0      0   1.1 4e-02    1.1   1.0   1.1  1.0   1 -0.21 -2.37
## cr       9   0      0   0.9 1e-01    0.9   0.8   0.9  0.7   1  0.08 -0.33
## plt      9   1     11 146.5 4e+01  156.0 126.0 162.8 91.0 204 -0.17 -0.08
## alb      9   1     11   3.9 5e-01    4.0   3.8   4.2  2.8   4 -1.55  2.34
## eag      9   0      0   0.6 5e-01    1.0   0.0   1.0  0.0   1 -0.27 -2.57
## dna_log  9   0      0   6.0 2e+00    6.0   4.8   7.6  1.0   9 -0.97  1.39
##
## p-values
##      pNormal pNonNormal
## age      0.2137236 0.5286397
## alt      0.5515067 0.3707812
## bil      0.6858601 0.5682000
## inr      0.5038224 0.8743690
## cr       0.7693120 0.6323041
## plt      0.5987924 0.3491690
## alb      0.2516633 0.1701735
## eag      0.6052414 0.5941235
## dna_log  0.8113600 0.6983231
##
## Standardize mean differences
##      1 vs 2
## age      0.5452206
## alt      0.2409274
## bil      0.1719907
## inr      0.3203851
## cr       0.1344336
## plt      0.2500112
## alb      0.5383001
## eag      0.2187767
## dna_log  0.1004008
##
```



```

## =====
##
##      ### Summary of categorical variables ###
##
## treat_gr: ETV
##      var  n miss p.miss level freq percent cum.percent
## gender 15    0   0.0    F    4   26.7      26.7
##              M   11   73.3      100.0
##
##      lc 15    0   0.0    0    6   40.0      40.0
##              1    9   60.0      100.0
##
## -----
## treat_gr: TDF
##      var  n miss p.miss level freq percent cum.percent
## gender  9    0   0.0    F    2   22.2      22.2
##              M    7   77.8      100.0
##
##      lc  9    0   0.0    0    4   44.4      44.4
##              1    5   55.6      100.0
##
##
## p-values
##      pApprox pExact
## gender      1      1
## lc          1      1
##
## Standardize mean differences
##      1 vs 2
## gender 0.10355607
## lc     0.09007547

```

```
write.csv(print(table1), file='C:\\Users\\phl02\\Desktop\\P\\bio\\ch2\\tableone.csv')
```

```
##          Stratified by treat_gr
##          ETV          TDF          p          test
##  n          15          9
##  age (mean (SD))  43.73 (9.16)  48.56 (8.52)  0.214
##  gender = M (%)    11 (73.3)    7 (77.8)  1.000
##  lc = 1 (%)        9 (60.0)    5 (55.6)  1.000
##  alt (mean (SD))  94.20 (93.09) 123.00 (141.11) 0.552
##  bil (mean (SD))   1.29 (0.41)   1.22 (0.42) 0.686
##  inr (mean (SD))   1.11 (0.19)   1.07 (0.04) 0.504
##  cr (mean (SD))    0.90 (0.20)   0.88 (0.11) 0.769
##  plt (mean (SD))  158.40 (56.88) 146.50 (36.00) 0.599
##  alb (mean (SD))   3.59 (0.64)   3.90 (0.52) 0.252
##  eag (mean (SD))   0.67 (0.49)   0.56 (0.53) 0.605
##  dna_log (mean (SD)) 5.78 (2.18)   6.01 (2.41) 0.811
```

11.3 gtsummary 패키지

```
library(gtsummary)

dat.temp<-dat %>%
  select(treat_gr,gender, age, lc, bil, inr, cr, dna_log)
```

기본 테이블

```
tbl_summary(dat.temp)
```

Characteristic	N = 24
treat_gr	
ETV	15 (63%)
TDF	9 (38%)
gender	
F	6 (25%)
M	18 (75%)
age	49 (43, 50)
lc	14 (58%)
bil	1.15 (1.00, 1.43)
inr	1.08 (1.03, 1.12)
Unknown	1
cr	0.91 (0.78, 1.00)
dna_log	6.17 (4.60, 7.52)

두 그룹으로 나누기

```
tbl_summary(dat.temp, by=treat_gr)
```

Characteristic	ETV, N = 15	TDF, N = 9
gender		
F	4 (27%)	2 (22%)
M	11 (73%)	7 (78%)
age	49 (37, 51)	49 (48, 50)
lc	9 (60%)	5 (56%)
bil	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)
inr	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)
Unknown	1	0
cr	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)
dna_log	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)

결측값 숨기기

```
tbl_summary(dat.temp, by=treat_gr, missing='no')
```

Characteristic	ETV, N = 15	TDF, N = 9
gender		
F	4 (27%)	2 (22%)
M	11 (73%)	7 (78%)
age	49 (37, 51)	49 (48, 50)
lc	9 (60%)	5 (56%)
bil	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)
inr	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)
cr	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)
dna_log	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)

결측값 다른 문구로 표시

```
tbl_summary(dat.temp, by='treat_gr', missing_text='missing value')
```

Characteristic	ETV, N = 15	TDF, N = 9
gender		
F	4 (27%)	2 (22%)
M	11 (73%)	7 (78%)
age	49 (37, 51)	49 (48, 50)
lc	9 (60%)	5 (56%)
bil	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)
inr	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)
missing value	1	0
cr	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)
dna_log	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)

평균과 표준편차로 나타내기

```
tbl_summary(dat.temp, by=treat_gr,  
             statistic=all_continuous()~ '{mean} \u00b1 {sd}',  
             missing='no')
```

Characteristic	ETV, N = 15	TDF, N = 9
gender		
F	4 (27%)	2 (22%)
M	11 (73%)	7 (78%)
age	44 ± 9	49 ± 9
lc	9 (60%)	5 (56%)
bil	1.29 ± 0.41	1.22 ± 0.42
inr	1.12 ± 0.19	1.07 ± 0.04
cr	0.90 ± 0.20	0.88 ± 0.11
dna_log	5.78 ± 2.18	6.01 ± 2.41

p값 표시

```
tbl_summary(dat.temp, by=treat_gr,
             missing='no') %>%
  add_p()
```

Characteristic	ETV, N = 15	TDF, N = 9	p-value
gender			>0.9
F	4 (27%)	2 (22%)	
M	11 (73%)	7 (78%)	
age	49 (37, 51)	49 (48, 50)	0.5
lc	9 (60%)	5 (56%)	>0.9
bil	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)	0.6
inr	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)	0.9
cr	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)	0.7
dna_log	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)	0.7

전체 N수 표시

```
tbl_summary(dat.temp, by=treat_gr,
             missing='no') %>%
  add_n()
```

Characteristic	N	ETV, N = 15	TDF, N = 9
gender	24		
F		4 (27%)	2 (22%)
M		11 (73%)	7 (78%)
age	24	49 (37, 51)	49 (48, 50)
lc	24	9 (60%)	5 (56%)
bil	24	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)
inr	23	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)
cr	24	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)
dna_log	24	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)

전체 환자 특성 같이 보여주기

```
tbl_summary(dat.temp, by=treat_gr,
             missing='no') %>%
  add_overall()
```

Characteristic	Overall, N = 24	ETV, N = 15	TDF, N = 9
gender			
F	6 (25%)	4 (27%)	2 (22%)
M	18 (75%)	11 (73%)	7 (78%)
age	49 (43, 50)	49 (37, 51)	49 (48, 50)
lc	14 (58%)	9 (60%)	5 (56%)
bil	1.15 (1.00, 1.43)	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)
inr	1.08 (1.03, 1.12)	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)
cr	0.91 (0.78, 1.00)	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)
dna_log	6.17 (4.60, 7.52)	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)

테이블 제목 변경

```
dat.temp %>%  
  tbl_summary(by=treat_gr,  
              missing='no') %>%  
  modify_caption('**Baseline Characteristics**')
```

Table 9: Baseline Characteristics

Characteristic	ETV, N = 15	TDF, N = 9
gender		
F	4 (27%)	2 (22%)
M	11 (73%)	7 (78%)
age	49 (37, 51)	49 (48, 50)
lc	9 (60%)	5 (56%)
bil	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)
inr	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)
cr	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)
dna_log	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)

테이블 소제목 변경

```
dat.temp %>%
  tbl_summary(by=treat_gr,
              missing='no') %>%
  modify_spanning_header( c('stat_1', 'stat_2')~'**Antiviral Treatment**')
```

Characteristic	ETV, N = 15	TDF, N = 9
gender		
F	4 (27%)	2 (22%)
M	11 (73%)	7 (78%)
age	49 (37, 51)	49 (48, 50)
lc	9 (60%)	5 (56%)
bil	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)
inr	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)
cr	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)
dna_log	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)

최종 출판용으로 정리된 테이블

```
tbl_summary(dat.temp,
            by=treat_gr,
            missing='no') %>%
  add_p() %>%
  add_overall() %>%
  modify_spanning_header( c('stat_1','stat_2')~'**Antiviral Treatment**') %>%
  modify_caption('**Baseline Characteristics**')
```

Table 11: Baseline Characteristics

Characteristic	Overall, N = 24	ETV, N = 15	TDF, N = 9	p-value
gender				>0.9
F	6 (25%)	4 (27%)	2 (22%)	
M	18 (75%)	11 (73%)	7 (78%)	
age	49 (43, 50)	49 (37, 51)	49 (48, 50)	0.5
lc	14 (58%)	9 (60%)	5 (56%)	>0.9
bil	1.15 (1.00, 1.43)	1.30 (1.05, 1.35)	1.10 (1.00, 1.50)	0.6
inr	1.08 (1.03, 1.12)	1.08 (0.99, 1.17)	1.10 (1.03, 1.10)	0.9
cr	0.91 (0.78, 1.00)	0.93 (0.71, 1.00)	0.90 (0.81, 0.91)	0.7
dna_log	6.17 (4.60, 7.52)	6.30 (3.71, 7.33)	6.04 (4.82, 7.56)	0.7