

```

1   ###第4回基本統計量,単純集計,量的データの Kategorize###
2   #量的・質的データの単純集計と量的データの Kategorize#
3   install.packages("psych")
4   install.packages("fdth")
5   dat <- read.csv("data/cars93.csv" ,
6                   row.names = 1,
7                   header = T )
8   head(dat)
9   #量的データ(間隔尺度,比例尺度)
10  library(psych)
11  describe(dat[c(5,7,8,12,13)]) #describe()関数:基本統計量
12  summary(dat[c(5,7,8,12,13)]) #summary()関数:最小値,四分位数,平均値,最大値
13  by(dat[c(5,7,8,12,13)] ,
14      dat$Origin ,
15      describe)                #by()関数:層別基本統計量
16                                #層別変数
17                                #適用関数
17  #質的データ(名義尺度,順序尺度)
18  table(dat$Type)                #単純集計(度数分布)
19  table(dat$Type:dat$Origin)    #Type×Origin単純集計
20
21  ###量的データを階級化し分布を確認###
22  library(fdth)
23  range(dat$EngineSize)         #EngineSizeのレンジを確認
24  L <- fdt(dat$EngineSize,      #fdt()関数:dataの階級化
25          start=1.0,           #最小値:小数点の場合は小数点以下を切り捨てた整数
26          end=6.0,             #最大値;小数点の場合は小数点以下を切り上げた整数
27          h=1)                  #階級幅:階級幅1なのでstart,endは上記設定が必要
28  L                              #結果を返す
29  ###階級化データの作成###
30  #right = F:区間の下限値は区間に含まれるが
31  #上限の値は区間に含まれない.TRUEは含まれる
32  ##上記Lの結果を参考に階級化する
33  cate.L <- cut( dat$EngineSize, #cut()関数
34                breaks = c(1,2,3,4,5,6), #区切り値
35                ##階級幅1以上-2未満,2以上-3未満,以下同##
36                labels= c("L","LM","M",
37                           "MH","H"),    #ラベル
38                right=F,                #上限値含まず
39                ordered_result = T)     #順序性
40  table(cate.L)                       #分布の確認
41  cate.L <- data.frame(cate.L)         #階級化をデータフレーム化
42  dat.c <- cbind(cate.L , dat)        #元dataにマージ
43  write.csv(dat.c,
44            "data/cars93.2.csv")      #2列目にcate.Lがマージ
45
46

```