

```

1   ###第7重回帰分析###
2   install.packages("glm2",dependencies = T)
3   install.packages("QuantPsyc",dependencies = T)
4   install.packages("DAAG",dependencies = T)
5   #install.packages("broom",dependencies = T)
6   dat <- read.csv("data/cars93.csv" ,
7                 header = T ,
8                 row.names = 1 )
9   head(dat , 3)
10  ###基本統計量,相関係数行列###
11  library(psych)
12  describe(dat[,c(4:8,12:15,17:23)])           #基本統計量NA有はcomplete.data数
13  round(cor(dat[,c("Price",                    #cor()関数:相関係数
14                "MPG.city",                    #round()関数:四捨五入
15                "EngineSize",
16                "Fuel.tank.capacity",          ##"Rear.seat.room"(NA有)
17                "Passengers")]),              # "p":積率相関,出力は小数第3位
18        method="p"),3)
19
20  pairs.panels(dat[,c("Price",                  #pairs.panels()関数
21                    "MPG.city",                ##NA有も出力
22                    "EngineSize",
23                    "Fuel.tank.capacity",
24                    "Passengers")]),
25        hist.col="red",                        #ヒストグラムの色
26        rug=F,ellipses=F,                      #ラグプロット,相関楕円
27        lm=T,density=F)                       #回帰直線,確率密度分布
28
29
30  ###重回帰分析### glm関数 or lm()関数:formula;y ~ x1 + x2 + x3 +...+ xn
31  library(glm2)
32  out<-glm(Price~MPG.city+EngineSize+          #glm()関数:一般線形回帰
33          Fuel.tank.capacity+Passengers,      #yの分布:正規分布
34          family = gaussian ,
35          data=dat)
36  summary(out)
37  anova(out)                                  #glm関数はanova出力なし
38  names(out)                                  #分析出力
39  out2<-lm(Price~MPG.city+                    #lm()関数:線形回帰
40          EngineSize+
41          Fuel.tank.capacity+
42          Passengers,
43          data=dat))
44  summary(out2)
45  names(out2)                                  #分析出力
46  out3<-lm(Price~MPG.city+                    #lm()関数:線形回帰(NAありdata)
47          EngineSize+
48          Fuel.tank.capacity+
49          Passengers,Rear.seat.room,          #Rear.seat.roomはNAあり
50          data=dat))
51  summary(out3)                                #NAがあるサンプル数出力
52  #多重共線性診断
53  library(DAAG)
54  vif(out2,digits = 2)                        #共線性診断:Package"car","HH"にもvif関数
55  #変数選択(AIC):引数無しは変数減少法
56  step(out2)
57  #変数選択:Step-Wise法の設定
58  out_1 <- lm(Price ~ 1 , data = dat)         #定数項だけのモデル作成
59  out_1
60  step(out_1 ,                                #定数項だけのモデル

```

```
61     scope = list(upper = ~MPG.city+           #scope:upper:全説明変数
62                   EngineSize+             #direction未指定:Stepwise
63                   Fuel.tank.capacity+
64                   Passengers,
65                   lower = ~ 1))             #lower:定数項だけモデル
66 ###残差###
67 Resi.m <- residuals(out2)
68 ###予測値###
69 Pred.m <- predict(out2)
70 dat <- data.frame(dat, Resi.m, Pred.m)       #データに残差、予測値をマージ
71 ###回帰診断###
72 library(ggplot2)
73 sindan <- fortify(out2)                     #cooks, fitted, resid, stdresidの出力
74 names(sindan)
75 head(sindan, 3)
76 par(mfrow=c(2, 2))                         #グラフを2行2列で表示
77 plot(out2)
78 par(mfrow=c(2, 3))                         #グラフを2行3列で表示
79 plot(out2, which = 1:6)
80 par(mfrow=c(1, 1))                         #グラフを1図表示に戻す
81
82 ##標準化偏回帰係数算出, lm.beta()関数###
83 library(QuantPsyc)
84 lm.beta(out2)                              #標準化偏回帰係数を計算
85 ###標準化データによる重回帰分析###
86 z <- scale(dat[, c("Price", "MPG.city",     #scale()関数:標準化
87                   "EngineSize",
88                   "Fuel.tank.capacity",
89                   "Passengers") 1])
90 z.dat <- data.frame(dat , z )              #元dataとZdataをデータフレーム化
91 names(z.dat)                              #項目名確認:Zは変数名.1になる
92 #write.csv(z.dat, "data/cars93.z.csv")    #Zdataをマージしたdataを保存
93 z.out <- (lm(Price.1~MPG.city.1+          #lm()関数
94             EngineSize.1+
95             Fuel.tank.capacity.1+
96             Passengers.1,
97             data=z.dat))
98 summary(z.out)
99
100
```