```r
1    ####二項ロジスティック回帰####################
2    ### (biopsy)data
3    library(MASS)
4    data("biopsy")
5    str(biopsy)           #Breast Cancer
6    colnames(biopsy)
7    table(biopsy$class)
8    #v1:clump thickness.塊の厚み
9    #v2:uniformity of cell size.細胞の大きさの均一性
10   #v3:uniformity of cell shape.細胞の形状の均一性
11   #v4:marginal adhesion.辺縁癒着
12   #v5:single epithelial cell size.単一の上皮細胞の大きさ
13   #v6:bare nuclei (16 values are missing).裸核(16個の値が欠損)
14   #v7:bland chromatin.(無菌の)染色質
15   #v8:normal nucleoli.正常な核小体
16   #v9:mitoses.有糸分裂
17   #calss:"benign" or "malignant".良性か悪性
18   #glm関数はfactor変数は自動的にDummy変数化して解析
19   #malignant=1,benign=0
20   bio.omit <- na.omit(biopsy)
21   library(ggplot2)
22   g2 <-   #重なりが多いのでjitterで作図
23       ggplot()+
24       geom_jitter(aes(x = V1, y = class, color = class), data = bio.omit)
25   g2
26
27   out.bi2 <- glm(class ~ V1 , data = bio.omit ,
28                  family = "binomial")
29   summary(out.bi2)
30   names(out.bi2)
31   #Null devianceとResidual devianceのカイ二乗検定
32   out.bi2.null <- glm(class ~ 1 , data = bio.omit ,
33                       family = "binomial")
34   anova(out.bi2.null , out.bi2 , test = "Chisq")
35
36   ori2 <- bio.omit[,c("class")]   #y
37   prd2 <- fitted.values(out.bi2) #確率
38   prd22 <- round(fitted.values(out.bi2)) #round関数で予測値を0,1で返す
39   d2 <- data.frame(ori2 , prd2 , prd22)
40   head(d2 , 10)
41   tab2 <- table(d2$ori2 , d2$prd22) #予測精度
42   tab2
43   round(((tab2[1,1]+
44           tab2[2,2])/sum(tab2)*100),2)        #的中率
45
46   ####説明変数を増やす
47   out.bi3 <- glm(class ~ V1 + V2 + V3 , data = bio.omit ,
48                  family = "binomial")
49   summary(out.bi3)
50   out.bi2.null <- glm(class ~ 1 , data = bio.omit ,
51                       family = "binomial")
52   anova(out.bi2.null , out.bi3 , test = "Chisq")
53   library(car)
54   vif(out.bi3)
55   ori22 <- bio.omit[,c("class")]   #y
56   prd22 <- fitted.values(out.bi3) #確率
57   prd23 <- round(fitted.values(out.bi3)) #round関数で予測値を0,1で返す
58   d22 <- data.frame(ori22 , prd22 , prd23)
59   head(d22 , 10)
60   tab22 <- table(d22$ori22 , d22$prd23)  #予測精度
61   tab22
```

```r
62    round(((tab22[1,1]+
63                tab22[2,2])/sum(tab22)*100),2)          #的中率
64    ###回帰診断
65    plot(influence(out.bi3)$hat)   #Leverage:てこ比
66    plot(cooks.distance(out.bi3)) #Cookの距離
67    influencePlot(out.bi3 , main = "Influence Plot:円のサイズはCookの距離への影響度")
68
69    ###変数選択:変数減少法,Step-Wise法###############################
70    step(out.bi3) #変数減少法
71    #Step-Wise法の設定
72    out.bi2.null <- glm(class ~ 1 , data = bio.omit ,   #切片だけのモデル
73                    family = "binomial")
74    step(out.bi2.null , direction = "both",           #direction = "both"
Step-Wise法
75        scope = (~V1+V2+V3+V4+V5+V6+V7+V8+V9))
76    ###最終Model
77    out.fin <- glm(class ~ V1 + V3 + V4 + V6 +V7 + V8 + V9 , data = bio.omit ,
78                family = "binomial")
79    summary(out.fin)
80
81    library(car)
82    vif(out.fin)
83    ori.y <- bio.omit[,c("class")]                  #y
84    prd.p <- fitted.values(out.fin)                 #確率
85    prd.01 <- round(fitted.values(out.fin))             #round関数で予測値を0,1で返す
86    d.fin <- data.frame(ori.y , prd.p , prd.01)
87    tab.fin <- table(d.fin$ori.y , d.fin$prd.01)         #予測精度
88    tab.fin
89    round(((tab.fin[1,1]+
90                tab.fin[2,2])/sum(tab.fin)*100),2)         #的中率
91    ##指数変換値の計算
92    exp(out.fin$coefficients)
93
94    ###回帰診断
95    plot(influence(out.fin)$hat)   #Leverage:てこ比
96    plot(cooks.distance(out.fin)) #Cookの距離
97    influencePlot(out.fin , main = "Influence Plot:円のサイズはCookの距離への影響度")
98
99    #######説明変数にfactor型(カテゴリカル変数)を取り込む#########################
100   library(MASS)
101   data(Cars93)
102   str(Cars93)
103   colnames(Cars93)
104   anyNA(Cars93[,c(1:27)])     #指定範囲の項目にNAがあるか
105   cars <- na.omit(Cars93)     #NAをすべてomitしたファイル作成
106   table(cars$Origin)
107   library(ggplot2)
108   g <-                 #重なりが多いのでjitterで作図
109       ggplot()+
110       geom_jitter(aes(x = EngineSize, y = Origin, color = Origin), data =
        cars)
111   g
112   library(glm2)
113   #glm関数はfactor変数は自動的にDummy変数化して解析
114   #non-USA=1,USA=0
115   out.bi <- glm(Origin ~ EngineSize + Price + DriveTrain, data = cars ,
116               family = "binomial")
117   summary(out.bi)
118   names(out.bi)
119
120   ori <- cars[,c(26)]    #y
```

```r
121    prd <- fitted.values(out.bi) #確率
122    prd.2 <- round(fitted.values(out.bi)) #round関数で予測値を0,1で返す
123    d <- data.frame(ori , prd , prd.2)
124    d
125    tab <- table(d$ori , d$prd.2) #予測精度
126    tab
127    round(((tab[1,1]+
128              tab[2,2])/sum(tab)*100),2)          #的中率
129
130    ############変数選択
131    step(out.bi)                        #変数減少法
132    out.bi.null.c <- glm(Origin ~ 1 , data = cars ,
133                    family = "binomial")
134    step(out.bi.null.c , direction = "both",       #Step-Wise法
135        scope = (~ Price+MPG.city+EngineSize+Horsepower+RPM+Fuel.tank.capacity+
136                  Width+Length+DriveTrain+AirBags))
137    ###最終Model
138    out.fin2 <- glm(Origin ~ Width + Price + Fuel.tank.capacity + EngineSize +
139                    DriveTrain , family = "binomial", data = cars)
140    summary(out.fin2)
141    library(car)
142    vif(out.fin2) #カテゴリカル変数があるので,GVIF^(1/(2*Df))をチェック
143    ori2.y <- cars[,c("Origin")]   #y
144    prd2.p <- fitted.values(out.fin2) #確率
145    prd2.01 <- round(fitted.values(out.fin2)) #round関数で予測値を0,1で返す
146    d2.fin <- data.frame(ori2.y , prd2.p , prd2.01)
147    tab2.fin <- table(d2.fin$ori2.y , d2.fin$prd2.01) #予測精度
148    tab2.fin
149    round(((tab2.fin[1,1]+
150              tab2.fin[2,2])/sum(tab2.fin)*100),2)          #的中率
151    ##指数変換値の計算
152    exp(out.fin2$coefficients)
153
```