

Statistinės duomenų analizės praktinės užduotys

2017

11. Logistinė regresija.

- (a) Edukologas tiria žargono vartojimą sąlygojančius veiksnius. Po smagaus humoro vakaro, kurio metu visi leipo juokais, sociologas apklause 22 žiūrovus. Kai kurie iš jų pasake, kad prisijuokė, kiti pasidžiaugė, kad prisižvengė. Žiūrovo linksmumas (kintamasis $Y = 0$ - žvengia, $Y = 1$ - juokiasi) amžius ir lytis (0 - vyras, 1 - moteris) pateikta lentelėje. Remdamiesi logistine regresija, nustatykite, ar galima pagal amžių atskirti besijuokiančius žiūrovus nuo žvengiančiųjų. Lytą panaudoti kaip pseudokintamąjį.

Linksmumas Y	1	0	1	0	0	1	1	1	1	1	0
Lytis x_1	1	1	1	1	1	1	1	1	1	1	0
Amžius x_2	64	56	40	24	24	40	56	64	40	40	32
Linksmumas Y	1	1	0	0	0	0	1	0	0	1	0
Lytis x_1	0	0	0	0	0	0	0	0	0	0	0
Amžius x_2	64	56	40	24	24	40	56	64	40	40	32

```
dat.a <- matrix(c(1, 1, 64, 0, 1, 56, 1, 1, 40, 0, 1, 24,
                  0, 1, 24, 1, 1, 40, 1, 1, 56, 1, 1, 64,
                  1, 1, 40, 1, 1, 40, 0, 0, 32, 1, 0, 64,
                  1, 0, 56, 0, 0, 40, 0, 0, 24, 0, 0, 24,
                  0, 0, 40, 1, 0, 56, 0, 0, 64, 0, 0, 40,
                  1, 0, 40, 0, 0, 32),
               byrow = TRUE,
               nrow = 22, ncol = 3)

dat.a<-as.data.frame(dat.a)
colnames(dat.a) <- c("Linksmumas", "Lytis", "Amzius")

modelis.a <- glm(Linksmumas~., family = "binomial", data = dat.a)
summary(modelis.a)

##
## Call:
## glm(formula = Linksmumas ~ ., family = "binomial", data = dat.a)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.22542  -0.71360  -0.02325   0.89152   1.72779
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.41527    2.33734  -2.317   0.0205 *
## Lytis        1.95555    1.17147   1.669   0.0951 .
## Amzius       0.10443    0.04715   2.215   0.0268 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 30.498  on 21  degrees of freedom
```

```
## Residual deviance: 20.495 on 19 degrees of freedom
## AIC: 26.495
##
## Number of Fisher Scoring iterations: 5
```

- (b) (ČM II.7.2) Ar galima pagal pajamas (PAJAMOS) ir darbo prestižiškumo indeksą (PREST) atpažinti, kad respondentas aukštąjį išsilavinimą turi (MOKSL=1) arba neturi (MOKSL=0)? Duomenys pateikti lentelėje.

PAJAMOS	3670	1923	3067	3811	3494	2012	1637	1265	2722
PREST	60	65	70	105	70	55	55	35	105
MOKSL	1	0	1	1	1	0	0	0	0
PAJAMOS	4050	1501	3340	3193	3125	4050	3458	2219	3781
PREST	135	50	65	60	95	115	65	65	90
MOKSL	1	0	1	1	0	1	0	0	1
PAJAMOS	2736	2568	3408	3298	3043	3536	3780	3798	
PREST	85	135	110	60	95	80	94	78	
MOKSL	0	0	0	1	1	1	1	1	

```
dat.b <- matrix(c(3670, 60, 1, 1923, 65, 0, 3067, 70, 1,
                 3811, 105, 1, 3494, 70, 1, 2012, 55, 0,
                 1637, 55, 0, 1265, 35, 0, 2722, 105, 0,
                 4050, 135, 1, 1501, 50, 0, 3340, 65, 1,
                 3193, 60, 1, 3125, 95, 0, 4050, 115, 1,
                 3458, 65, 0, 2219, 65, 0, 3781, 90, 1,
                 2736, 85, 0, 2568, 135, 0, 3408, 110, 0,
                 3298, 60, 1, 3043, 95, 1, 3536, 80, 1,
                 3780, 94, 1, 3798, 78, 1),
               byrow = TRUE,
               nrow = 26, ncol = 3)

dat.b<-as.data.frame(dat.b)
colnames(dat.b) <- c("Pajamos", "Prest", "Moksl")

modelis.b <- glm(Moksl~., family = "binomial", data = dat.b)
summary(modelis.b)

##
## Call:
## glm(formula = Moksl ~ ., family = "binomial", data = dat.b)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3142  -0.1646   0.2124   0.4382   1.5880
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -11.415473    5.921251  -1.928  0.0539 .
## Pajamos      0.004976    0.001958   2.541  0.0110 *
## Prest       -0.048978    0.037076  -1.321  0.1865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 35.890 on 25 degrees of freedom
## Residual deviance: 13.877 on 23 degrees of freedom
## AIC: 19.877
##
## Number of Fisher Scoring iterations: 6
```

(c) Gimdymo namuose surinkti duomenys apie gimdyvių, svorį (kg), rūkymą (1 - rūko, 0 - nerūko), hipertonią (1 - serga, 0 - neserga) ir naujagimio svorį (g)

Motinos amžius	Rūkymas	Hipertonija	Naujag. svoris (g)	Motinos svoris (kg)
24	0	0	1703	64,0
21	1	1	1792	82,5
21	0	0	1930	100,0
19	0	0	2084	51,0
24	0	0	2102	69,0
17	1	0	2227	55,0
18	0	0	2284	74,0
15	0	0	2383	57,5
17	0	0	2440	60,0
20	0	0	2452	52,5
14	1	0	2468	50,5
14	0	0	2497	50,0
21	1	1	2497	65,0
33	0	0	2553	77,5
32	0	0	2837	60,5
28	0	0	2879	83,5
29	0	0	2922	75,0
26	1	0	2922	84,0
17	0	0	2922	56,5
35	1	0	2950	60,5
33	1	0	3035	54,5
21	1	0	3044	92,5
19	0	0	3064	94,5
21	0	0	3064	80,0
19	0	0	3177	57,5
28	0	0	3236	70,0
16	1	0	3376	67,5
22	0	0	3462	65,5
32	0	0	3475	85,0
19	0	0	3574	52,5
24	0	0	3730	55,0
25	0	1	3985	60,0

```
dat.c <- matrix(c(24, 0, 0, 1703, 64.0, 21, 1, 1, 1792, 82.5,
                  21, 0, 0, 1930, 100.0, 19, 0, 0, 2084, 51.0,
                  24, 0, 0, 2102, 69.0, 17, 1, 0, 2227, 55.0,
                  18, 0, 0, 2284, 74.0, 15, 0, 0, 2383, 57.5,
                  17, 0, 0, 2440, 60.0, 20, 0, 0, 2452, 52.5,
                  14, 1, 0, 2468, 50.5, 14, 0, 0, 2497, 50.0,
                  21, 1, 1, 2497, 65.0, 33, 0, 0, 2553, 77.5,
                  32, 0, 0, 2837, 60.5, 28, 0, 0, 2879, 83.5,
                  29, 0, 0, 2922, 75.0, 26, 1, 0, 2922, 84.0,
```

```

17, 0, 0, 2922, 56.5, 35, 1, 0, 2950, 60.5,
33, 1, 0, 3035, 54.5, 21, 1, 0, 3044, 92.5,
19, 0, 0, 3064, 94.5, 21, 0, 0, 3064, 80.0,
19, 0, 0, 3177, 57.5, 28, 0, 0, 3236, 70.0,
16, 1, 0, 3376, 67.5, 22, 0, 0, 3462, 65.5,
32, 0, 0, 3475, 85.0, 19, 0, 0, 3574, 52.5,
24, 0, 0, 3730, 55.0, 25, 0, 1, 3985, 60.0),
byrow = TRUE,
nrow = 32, ncol = 5)

dat.c<-as.data.frame(dat.c)
colnames(dat.c) <- c("Motinos_amzius", "Rukymas", "Hipertonija",
                    "Naujag_svoris", "Motinos_svoris")

```

Naujagimis sveria nepakankamai, jeigu jo svoris nesiekia 2500 g. Įvertinkite tikimybę, kad 38 metų būsimoji motina, kuri rūko, serga hipertonią ir sveria 85kg, pagimdys nepakankamo svorio naujagimį. Kokį naujagimio svorį prognozuotumėte taikydami tiesinę regresiją?

```

dat.c$y <- ifelse(dat.c$Naujag_svoris < 2500, 1, 0)
modelis.c <- glm(y~Motinos_amzius+Rukymas+Hipertonija+Motinos_svoris,
                family = "binomial", data = dat.c)

ats <- data.frame(38, 1, 1, 85)
colnames(ats)<-c("Motinos_amzius", "Rukymas",
                "Hipertonija","Motinos_svoris")

predict(modelis.c, ats, type = "response")

```

```

##          1
## 0.01368109

```

```
summary(modelis.c)
```

```

##
## Call:
## glm(formula = y ~ Motinos_amzius + Rukymas + Hipertonija + Motinos_svoris,
##      family = "binomial", data = dat.c)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5825  -0.8384  -0.1915   0.8284   1.7612
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   6.548689   2.956603   2.215   0.0268 *
## Motinos_amzius -0.303082   0.135700  -2.233   0.0255 *
## Rukymas        -0.326530   1.138149  -0.287   0.7742
## Hipertonija     1.740476   1.485841   1.171   0.2414
## Motinos_svoris -0.008512   0.031939  -0.266   0.7899
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 43.230  on 31  degrees of freedom

```

```
## Residual deviance: 30.598  on 27  degrees of freedom
## AIC: 40.598
##
## Number of Fisher Scoring iterations: 5
```

Padaryta su R version 3.4.2 (2017-09-28), x86_64-pc-linux-gnu.