Project

Simphiwe Mngadi

10 May 2023

Luekemia patients

Description

Christensen (1997) provides data on survival of 33 leukemia patients along with two potential predictor variables: white blood cell count (measured in thousands of cells per mm3 of blood) and the existence of a certain morphological characteristic in the cells called AG. If the characteristic is present the patient is said to be 'AG positive' and if it is absent the patient is said to be 'AG negative.' The dependent variable is yi as defined below, and the independent variables are xi1 (white blood cell count) and xi2 as defined below.

yi = {1 if ith patient survived for at least 52 weeks 0 if ith patient died within 52 weeks

xi2 is a binary explanatory variable defined as follows: xi2 = {1 if the ith patient is AG positive 0 if the ith patient is AG negative.

fitting the logistic model

```
leukdat <- read.csv("leukemia.csv", header = TRUE)
mylogistic <- glm(Survival ~ CellCount + AG, data = leukdat,
family = binomial(link = logit))
summary(mylogistic)
```

```
##
## Call:
## glm(formula = Survival ~ CellCount + AG, family = binomial(link = logit),
##     data = leukdat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5224  -0.6417  -0.4534   0.8362   2.1570
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.30734    0.81448  -1.605   0.1085
## CellCount   -0.03177    0.01863  -1.705   0.0881 .
## AG           2.26107    0.95222   2.375   0.0176 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 42.010  on 32  degrees of freedom
## Residual deviance: 31.062  on 30  degrees of freedom
```

```
## AIC: 37.062
##
## Number of Fisher Scoring iterations: 5
```

Statistical significance of coefficients

B2 (AG) is statistically significant at 5% level since the p-value for its significance test is 0.01757 $<$ 0.05 . B1 (CellCount) is not statistically significant since the corresponding p-value is 0.08814 $>$ 0.05.

Parameter estimates in practical terms

exp{B1} = exp{3.180 x 10^-2} = 1.032 . This implies that, if white blood cell count increases by 1000 cells per mm3 of blood , the expected odds of a patient surviving leukemia for 52 weeks decrease by a factor of 1.032. (Or, they change by a factor of 0.9687) .

exp{B2} = exp{2.261} =9.593 . Expected odds of survival for 52 weeks for a leukemia patient whose blood is AG positive are about 9.593 times as much as expected odds of survival for 52 weeks for a leukemia patient whose blood is AG negative .

Exposure to chloracetic acid and the death of mice

Description

Christensen (1997) provides data from experiments conducted to examine the relationship between exposure to chloracetic acid and the death of mice. Twelve different dose levels were used and ten mice were exposed at each dose level, yielding a total of 120 observations. Dose level is measured in grams per kg of body weight.

The dependent variable is defined as follows.

yi= {1 if ith mouse died 0 if ith mouse did not die

fit the probit regression model

```r
aciddat <- read.csv("chloracetic_acid.csv", header = TRUE)
myprobit <- glm(Died ~ Dose, data = aciddat,
family = binomial(link = probit))
summary(myprobit)
```

```
##
## Call:
## glm(formula = Died ~ Dose, family = binomial(link = probit),
##     data = aciddat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6326  -0.7198  -0.5720   0.9618   2.2758
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.1337     0.3943  -5.411 6.25e-08 ***
## Dose          8.7460     1.9068   4.587 4.50e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 151.34  on 119  degrees of freedom
## Residual deviance: 127.92  on 118  degrees of freedom
## AIC: 131.92
##
## Number of Fisher Scoring iterations: 4
```

sign of the coefficient estimate B1 and explain what this means for the dose-response relationship.

The sign of B1 is positive and B1 is statistically significant at 5% level of significance (p-value is $4.5 \times 10\hat{}6 < 0.05$) so we can see that as the dose of chloracetic acid increases, the probability of death for the mouse also increases .

LD50 and explain what it means.

LD50= -B0/B1= 2.133656/8.745963 = 0.2440. We estimate that a dose of 0.2440 g/kg of chloracetic acid is sufficient to kill a mouse with 50% probability.

Phat for each of the twelve dose levels represented in the data

```
phat <- unique(myprobit$fitted.values)
yhat <- as.integer(unique(myprobit$fitted.values) >= 0.5)
cbind(phat, yhat)
```

```
##             phat yhat
##  [1,] 0.07504319    0
##  [2,] 0.10400440    0
##  [3,] 0.15090981    0
##  [4,] 0.18463241    0
##  [5,] 0.20560634    0
##  [6,] 0.22819723    0
##  [7,] 0.28142156    0
##  [8,] 0.34869863    0
##  [9,] 0.43036901    0
## [10,] 0.52524814    1
## [11,] 0.62966132    1
## [12,] 0.73624696    1
```

which dose levels would we predict that a mouse will die, using a threshold value of 0.5?

It would be for dose levels 0.2512, 0.2818, and 0.3162, since these are the dose levels for which phat > 0.5 .