A2 Project 2

Johnsen & Johnsen

2022-10-18

Projekt 2: Survival Data

Analysis of the Binary Data

Read the data Logistic.txt into R.

```
# if (Sys.getenv("LOGNAME") == "mortenjohnsen"){
#   setwd("/Users/mortenjohnsen/OneDrive - Danmarks Tekniske Universitet/DTU/9. Semester/02418 - Statis
# } else {
#   setwd("~/Documents/02418 Statistical Modelling/Assignments/Assignment 1/Project-1")
# }

log.data <- read.table("Logistic.txt", header=TRUE, sep="",
                       as.is=TRUE)

source("testDistribution.R")
```

Fit a logistic regression model for the binary outcome AIDS=”yes” versus AIDS=”no” with the explanatory variable treatment with AZT (Yes, NO). Present the odds ratio for the effect of AZT on AIDS with 95% confidence interval and interpret the result in words

The logistic regression model is given by the likelihood function:

$$L(\theta) = \prod_i \left(\frac{\theta_i}{1-\theta_i}\right)^{y_i} (1-\theta_i)$$

Here θ is calculated as:

$$\theta_i = \frac{e^{\beta_0+\beta_1 t_{AZT}}}{1+e^{\beta_0+\beta_1 t_{AZT}}}$$

```
nll.log <- function(theta, event, n, treatment){

  beta0 <- theta[1]
  beta1 <- theta[2]

  theta <- exp(beta0 + beta1 * treatment)/(1 + exp(beta0 + beta1 * treatment))

  nll <- -sum(log(theta)*event  + log(1-theta)*(n - event))
  return(nll)
```

```
}

# For AZT treatment:
log.reg <- nlminb(start = c(1,1), objective = nll.log, event = log.data$AIDS_yes, n = log.data$n, treatr

beta0 <- log.reg$par[1]
beta1 <- log.reg$par[2]
```

```
logistic <- data.frame("AZT" = c(rep(1,170), rep(0,168))
                       ,"AIDS_yes" = c(rep(c(1,0),c(25,170-25)), rep(c(1,0), c(44, 168-44))))

fit.glm <- glm(AIDS_yes ~ AZT, data = logistic, family = binomial)
print(cat(paste0("with glm model: ", coef(fit.glm)
  ,"\nBy hand (according to slide 19 lect 4): "
  ,"\nbeta_0 = ", log.reg$par[1], ", beta_1 = ", log.reg$par[2])))
```

```
## with glm model: -1.03609193168383
## By hand (according to slide 19 lect 4):
## beta_0 = -1.03609192753506, beta_1 = -0.721765976832429 with glm model: -0.721765985868547
## By hand (according to slide 19 lect 4):
## beta_0 = -1.03609192753506, beta_1 = -0.721765976832429NULL
```

Calculating 95% wald confidence interval for β_1 and subsequently for the odds ratio.

```
sd <- sqrt(diag(solve(hessian(func = nll.log, x = c(beta0, beta1), event = log.data$AIDS_yes, n = log.da

beta1.wald.CI <- beta1 + c(-1,1)*dnorm(0.975)*sd[2]
```

Odds ratio $= exp(\beta_1) = 0.4858934$, $[0.4534386, 0.5206711]$. Thus for the individuals receiving the AZT treatment, the odds of developing AIDS or dying is reduced by a factor 0.486.

Test the hypothesis of no effect of AZT on AIDS using:

- The likelihood ratio test
- The Wald test
- The score test

Likelihood ratio test

Analysis of the survival time data

Descriptive statistics

Read the data actg320.txt into R

How many patients got AIDS or died in the two treatment groups? And how long was the total follow-up time in the two groups?

Plot the survival functions in the two treatment groups, which group seems to be doing best?

Plot the cumulative incidence functions for the two groups, which plot would you prefer?

Compare the survival in the two treatment groups using a log-rank test.

Parametric survival models

Fit parametric survival models containing treatment (tx) and CD4 count (cd4) as explanatory variables

- Try using the exponential, Weibull and log-logistic models, which one gave the best fit (and why)?

Using the survival model you chose, make a table of estimates and their 95% confidence intervals

Using your model compute the time ratio for the treatment effect. Sim- ilarly, compute the time ratio for the effect of increasing the CD4 count with 50. In both cases unceartainty evaluation (e.g. confidence inter- vals) should be included. Interpret the results in words

Assess the goodness of fit of this model using a plot based on the Cox Snell residuals

Give a graphical presentation of your model