

Laboratorium 6

1. Regresja logistyczna

W pewnym szpitalu przeprowadzono badanie skuteczności pewnego testu medycznego na obecność pewnej choroby. Zgromadzono 47 pacjentów chorych na tę chorobę oraz 79 pacjentów, u których wykluczono tę chorobę. Następnie każdy z pacjentów został poddany testowi medycznemu którego duże wartości świadczą o chorobie. Wyniki przedstawione są w pliku Dane 4.1. Korzystając z modelu regresji logistycznej wyznacz funkcję szacującą prawdopodobieństwo wystąpienia choroby na podstawie wyniku testu medycznego oraz przetestuj hipotezę o braku zależności między wynikiem testu medycznego a wystąpieniem choroby.

```
library(readxl)

library(caret)

Dane <- read_excel("Dane 4.1.xlsx")

D<-Dane$"D"

T<-Dane$"T"

n<-length(D)

lik<-function(x)
{
  x1 <- x[1]
  x2 <- x[2]
  L<-D*log(1/(1+exp(-(x1+x2*T))))+(1-D)*log(1-1/(1+exp(-(x1+x2*T))))
  return(-sum(L))
}

Opt<-optim(c(0,0), lik)

C<-Opt$par

TS<-sort(T)

P<-1/(1+exp(-(C[1]+C[2]*TS)))

plot(T,D)

lines(TS,P,col="blue")

C

V<-cbind(T,1/(1+exp(-(C[1]+C[2]*T))),D)

V
```

```
likH0<-function(x)
{
  L<-D*log(1/(1+exp(-(x))))+(1-D)*log(1-1/(1+exp(-(x))))
  return(-sum(L))
}
```

```
OptH0<-optim(0, likH0, method = "SANN")
```

```
C0<-OptH0$par
```

```
2*(likH0(C0)-lik(C))
```

```
Test<-2*(-lik(C)+likH0(C0))
```

```
pval<-1-pchisq(Test,1)
```

```
pval
```

Sprawdzimy teraz empirycznie czy przy H_0 statystyka testowa ma rozkład chi-kwadrat o 1 stopniu swobody:

```
TestH0<-c()
```

```
pvalH0<-c()
```

```
for(k in 1:1000)
```

```
{
```

```
  TMC<-runif(n,0,100)
```

```
  likMC<-function(x)
```

```
  {
```

```
    x1 <- x[1]
```

```
    x2 <- x[2]
```

```
    L<-D*log(1/(1+exp(-(x1+x2*TMC))))+(1-D)*log(1-1/(1+exp(-(x1+x2*TMC))))
```

```
    return(-sum(L))
```

```
  }
```

```
  OptMC<-optim(c(0,0), likMC)
```

```
  CMC<-OptMC$par
```

```
likH0MC<-function(x)
```

```
{
```

```
  L<-D*log(1/(1+exp(-(x))))+(1-D)*log(1-1/(1+exp(-(x))))
```

```
  return(-sum(L))
```

```

}

OptH0MC<-optim(0, likH0MC, method = "SANN")

C0MC<-OptH0MC$par

TestH0[k]<-2*(-likMC(CMC)+likH0MC(C0MC))

pvalH0[k]<-1-pchisq(TestH0[k],1)

}

plot(ecdf(pvalH0))

plot(ecdf(TestH0))

x<-10*(c(1:n)-1)/n

lines(x,pchisq(x,1),col="blue")

```

Gotowiec do estymacji i testu istotności:

```

model <- glm( D ~ T,
              data = Dane, family = binomial)

summary(model)$coef

C<-model$coefficients

TS<-sort(T)

P<-1/(1+exp(-(C[1]+C[2]*TS)))

plot(T,D)

lines(TS,P,col="blue")

```