STAT 2210 - Homework 4

Giang Vu

10/13/2021

1. Problem 1

(a) Equation for logistic regression model

$$ln(\frac{\hat{\pi}}{1-\hat{\pi}}) = -3.77714 + 0.14486(LI)$$

(b) Estimated probabilities of remission

When $LI = 8$,

$$\hat{\pi}(8) = \frac{e^{-3.77714+0.14486(8)}}{1+e^{-3.77714+0.14486(8)}} = 0.06797244$$

When $LI = 26$,

$$\hat{\pi}(26) = \frac{e^{-3.77714+0.14486(26)}}{1+e^{-3.77714+0.14486(26)}} = 0.497305$$

When $LI = 34$,

$$\hat{\pi}(34) = \frac{e^{-3.77714+0.14486(34)}}{1+e^{-3.77714+0.14486(34)}} = 0.7591637$$

(c) Rate of change in $\hat{\pi}$ when $LI = 26$

$$\hat{\beta}_1\hat{\pi}(26)(1-\hat{\pi}(26)) = 0.14486(0.497305)(1-0.497305) = 0.036$$

(d) Change in $\hat{\pi}$ from $LI = 14$ to $LI = 28$

When $LI = 14$,

$$\hat{\pi}(14) = \frac{e^{-3.77714+0.14486(14)}}{1+e^{-3.77714+0.14486(14)}} = 0.148$$

When $LI = 28$,

$$\hat{\pi}(28) = \frac{e^{-3.77714+0.14486(28)}}{1+e^{-3.77714+0.14486(28)}} = 0.569$$

Therefore,

$$\hat{\pi}(28) - \hat{\pi}(14) = 0.57 - 0.15 = 0.42$$

(e) Change in estimated odds of remission with a one unit increase in LI

$$e^{\hat{\beta}_1} = e^{0.14486} = 1.16$$

(f) 95% confidence interval for π of remission when $LI = 26$

The 95% CI for $\beta_0 + \beta_1 x$ when $LI = 26$ is

$$\hat{\pi}(26) \pm 1.96\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1(26))}$$

We have

$$Var(\hat{\beta}_0 + \hat{\beta}_1(26)) = Var(\hat{\beta}_0) + 2(26)Cov(\hat{\beta}_0, \hat{\beta}_1) + 26^2 Var(\hat{\beta}_1) = 1.9 - 2(26)(0.077) + 26^2(0.004) = 0.6$$

Therefore, the 95% CI for $\beta_0 + \beta_1 x$ when $LI = 26$ is

$$0.497305 \pm 1.96\sqrt{0.6}$$

which is $(-1.021, 2.016)$.
So, the 95% for π when $LI = 26$ is $(0.265, 0.882)$.

(g) Wald test for the LI effect at $\alpha = 0.05$

Null hypothesis: $H_0 : \beta_1 = 0$
Alternative hypothesis: $H_A : \beta_1 \neq 0$
Test statistic is

$$Z = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{0.145}{0.06} = 2.417$$

which follows the standard normal distribution. Thus the p-value is

$$2P(Z \geq 2.417) = 2(1 - 0.992) = 0.015 < \alpha$$

Therefore, we reject the null hypothesis and conclude that we have strong evidence of a positive effect of LI on the cancer remission probability.

(h) 95% Wald CI for the odds ratio corresponding to a 1-unit increase in LI

The odds ratio is e^{β_1}, so we need to construct 95% Wald CI for log odds ratio first, which is β_1.
The 95% Wald CI for $beta_1$ is $0.145 \pm 1.96(0.06)$ which will be $(0.0274, 0.2626)$.
Therefore, the 95% Wald CI for the odds ratio is $(1.03, 1.30)$.
Interpretation: We are 95% confident that a 1-unit increase in LI has at least 3% increase and at most 30% increase in the odds that a patient achieves cancer remission.

(i) Likelihood ratio test for the LI effect at $\alpha = 0.05$

Null hypothesis: $H_0 : \beta_1 = 0$
Alternative hypothesis: $H_A : \beta_1 \neq 0$
The test statistic is: $-2(l_0 - l_1) = 34.372 - 26.073 = 8.299$, which follows a Chi-squared distribution with 1 degree of freedom.
The p-value is $P(\chi^2 \geq 8.299) = 0.004 < \alpha$

Therefore, we reject the null hypothesis and conclude that we have strong evidence of a positive effect of LI on the cancer remission probability.

(j) 95% likelihood ratio CI for the odds ratio

The 95% likelihood ratio CI for the log odds ratio, β_1 is given in the R output, which is $(0.043, 0.285)$.
Then the 95% likelihood ratio CI for the odds ratio, e^{β_1}, is $(1.04, 1.33)$.
Interpretation: We are 95% confident that a 1-unit increase in LI has at least 4% increase and at most 33% increase in the odds that a patient achieves cancer remission.

2. Problem 2

(a) Equation for logistic regression model for π = probability of a satellite

$$ln(\frac{\hat{\pi}}{1-\hat{\pi}}) = -3.6947 + 1.8151(weight)$$

```r
crabs <- read.table("Crabs.txt", header = T)
crabs.fit <- glm(y ~ weight, family = binomial(link=logit), data = crabs)
crabs.summary <- summary(crabs.fit)
crabs.summary
```

```
##
## Call:
## glm(formula = y ~ weight, family = binomial(link = logit), data = crabs)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1108  -1.0749   0.5426   0.9122   1.6285
##
## Coefficients:
##             Estimate Std. Error z value   Pr(>|z|)
## (Intercept)  -3.6947     0.8802  -4.198 0.00002697 ***
## weight        1.8151     0.3767   4.819 0.00000145 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 225.76  on 172  degrees of freedom
## Residual deviance: 195.74  on 171  degrees of freedom
## AIC: 199.74
##
## Number of Fisher Scoring iterations: 4
```

(b) Estimated probabilities of having a satellite

When $weight = 1.20$,

$$\hat{\pi}(1.20) = \frac{e^{-3.6947+1.8151(1.20)}}{1 + e^{-3.6947+1.8151(1.20)}} = 0.180$$

When $weight = 2.44$,

$$\hat{\pi}(2.44) = \frac{e^{-3.6947+1.8151(2.44)}}{1 + e^{-3.6947+1.8151(2.44)}} = 0.675$$

When $weight = 5.20$,

$$\hat{\pi}(5.20) = \frac{e^{-3.6947+1.8151(5.20)}}{1 + e^{-3.6947+1.8151(5.20)}} = 0.997$$

```r
b0 <- coef(crabs.fit)[1]
b1 <- coef(crabs.fit)[2]
#predict pi for weight = 1.2
pi.min <- exp(b0+b1*min(crabs$weight)) / (1+ exp(b0+b1*min(crabs$weight)))
pi.min
```

```
## (Intercept)
##   0.1799697
```

```r
#predict pi for weight = 2.44
pi.avg <- exp(b0+b1*mean(crabs$weight)) / (1+ exp(b0+b1*mean(crabs$weight)))
pi.avg
```

```
## (Intercept)
##   0.6746137
```

```r
#predict pi for weight = 5.2
pi.max <- exp(b0+b1*max(crabs$weight)) / (1+ exp(b0+b1*max(crabs$weight)))
pi.max
```

```
## (Intercept)
##   0.9968084
```

(c) Rate of change in $\hat{\pi}$ when $x = 3.0$

$$\hat{\beta}_1 \hat{\pi}(3)(1 - \hat{\pi}(3)) = 0.229$$

```r
pi.3 <- exp(b0+b1*3) / (1+ exp(b0+b1*3))
b1*pi.3*(1-pi.3)
```

```
##    weight
## 0.2288288
```

(d) Weight when $\hat{\pi} = 0.50$

At the median effective level, the weight is $x = -\frac{\hat{\beta}_0}{\hat{\beta}_1} = -\frac{-3.6947}{1.8151} = 2.0355$ kg.

```
-(b0/b1)
```

```
## (Intercept)
##      2.0355
```

(e) 95% Wald CI for the odds ratio corresponding to a 1-unit increase in weight

The odds ratio is e^{β_1}, so we need to construct 95% Wald CI for log odds ratio first, which is β_1.
The 95% Wald CI for $beta_1$ is $1.8151 \pm 1.96(0.3767)$ which will be $(1.077, 2.553)$.
Therefore, the 95% Wald CI for the odds ratio is $(2.935, 12.851)$.
Interpretation: We're 95% confident that a 1-unit increase in weight has at least 193% increase and at most 1185% increase in the odds that a female crab has a satellite.

```
data.frame("LB"=b1-qnorm(0.975)*crabs.summary$coefficients[2,2],
           "UB"=b1+qnorm(0.975)*crabs.summary$coefficients[2,2],
           row.names = "Wald CI for log odds")
```

```
##                            LB       UB
## Wald CI for log odds 1.076834 2.553455
```

```
data.frame("LB"=exp(b1-qnorm(0.975)*crabs.summary$coefficients[2,2]),
           "UB"=exp(b1+qnorm(0.975)*crabs.summary$coefficients[2,2]),
           row.names = "Wald CI for odds")
```

```
##                        LB       UB
## Wald CI for odds 2.935372 12.85143
```

(f) Wald test for the weight effect at $\alpha = 0.05$

Null hypothesis: $H_0 : \beta_1 = 0$
Alternative hypothesis: $H_A : \beta_1 \neq 0$
Test statistic is

$$Z = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{1.815}{0.377} = 4.819$$

which follows the standard normal distribution. Thus the p-value is

$$2P(Z \geq 4.819) = 0.000001445736 < \alpha$$

Therefore, we reject the null hypothesis and conclude that we have strong evidence of a positive effect of weight on the satellite probability of a female crab.

```
z <- b1/(crabs.summary$coefficients[2,2])
(1-pnorm(z))*2
```

```
##          weight
## 0.000001445736
```

(g) Likelihood ratio test for the LI effect at $\alpha = 0.05$

Null hypothesis: $H_0 : \beta_1 = 0$
Alternative hypothesis: $H_A : \beta_1 \neq 0$
The test statistic is: $-2(l_0 - l_1) = 225.76 - 195.74 = 30.02$, which follows a Chi-squared distribution with 1 degree of freedom.
The p-value is $P(\chi^2 \geq 30.02) = 0.00000004273103 < \alpha$

```
chi_sqr <- crabs.summary$null.deviance - crabs.summary$deviance
1-pchisq(chi_sqr, df=1)
```

```
## [1] 0.00000004273103
```

Therefore, we reject the null hypothesis and conclude that we have strong evidence of a positive effect of weight on the satellite probability of a female crab.

(h) 95% likelihood ratio CI for the odds ratio

The 95% likelihood ratio CI for the log odds ratio, β_1 is given in the R output, which is $(1.114, 2.597)$.
Then the 95% likelihood ratio CI for the odds ratio, e^{β_1}, is $(3.05, 13.43)$.
Interpretation: We're 95% confident that a 1-unit increase in weight has at least 204.59% increase and at most 1242.75% increase in the odds that a female crab has a satellite.

```
LR.ci <- confint(crabs.fit)
```

```
## Waiting for profiling to be done...
```

```
LR.lo <- data.frame("LB"=LR.ci[2,1],
           "UB"=LR.ci[2,2],
           row.names = "Likelihood ratio CI for log odds")
LR.lo
```

```
##                                       LB       UB
## Likelihood ratio CI for log odds 1.11379 2.597305
```

```
LR.o <- data.frame("LB"=exp(LR.ci[2,1]),
           "UB"=exp(LR.ci[2,2]),
           row.names = "Likelihood ratio CI for odds")
LR.o
```

```
##                                   LB      UB
## Likelihood ratio CI for odds 3.04588 13.4275
```