Homework five

Yi Chen(yc3356)

February 26, 2018

Homework five

```
# read the data
setwd("C:/Users/cheny/Desktop/study/second term/Advanced Data Analysis/homework/homework five")
adolesent <- read.csv("adolescent.csv")
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'adolescent.csv'
```

```
shuttle <- read.csv("Shuttle.csv")
```

problem one

(a)

Use logistic regression to model the effect of the temperature on the probability of thermal distress.

```
fit1 <- glm(factor(shuttle$ThermalDistress)~shuttle$Temperature,family = binomial("logit"))
summary(fit1)
```

```
##
## Call:
## glm(formula = factor(shuttle$ThermalDistress) ~ shuttle$Temperature,
##     family = binomial("logit"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0611  -0.7613  -0.3783   0.4524   2.2175
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          15.0429     7.3786   2.039   0.0415 *
## shuttle$Temperature  -0.2322     0.1082  -2.145   0.0320 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.267  on 22  degrees of freedom
## Residual deviance: 20.315  on 21  degrees of freedom
## AIC: 24.315
##
## Number of Fisher Scoring iterations: 5
```

```
exp(coef(fit1))
```

```
##         (Intercept) shuttle$Temperature
##        3.412315e+06        7.928171e-01
```

analysis

As we can see from the table, the result of the fitted regression is:

$$logit(\pi(TD|Temperature)) = 15.0429 - 0.2322Temperature$$

(b)

Estimate $\beta1$, the effect of temperature on the probability of thermal distress. Interpret your result.

Analysis

As we can see from the result: the estimated value of $\beta_1 = -0.2322$. This indecate that on average, if we ignore the effect of other underlying factors, if the temperature increase 1 unit(1 fahrenheit), the the odd of at least one primary O-ring suffered termal distress would be $e^{-0.2322} = 0.7927875$ times as before. i.e. will have approximately 20.72% of decrease.

(C)

Construct a 95% confidence interval to describe the effect of the temperature on the odds of thermal distress, Interpret your result.

```
confint(fit1)
```

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

```
##                            2.5 %      97.5 %
## (Intercept)           3.3305848 34.34215133
## shuttle$Temperature -0.5154718 -0.06082076
```

```
exp(confint(fit1))
```

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

```
##                            2.5 %       97.5 %
## (Intercept)          27.9546841 8.214986e+14
## shuttle$Temperature  0.5972188 9.409919e-01
```

Analysis

As we can see from the table, the confidence interval of β_1 is between -0.5154718 and -0.06082076. This indecate that, on average, if we ignore the effect of other underlying factors, we have 95 percent of confidenct to conclude that if the temperature increase 1 unit(1 fahrenheit), the the odd of at least one primary O-ring suffered termal distress would between $e^{-0.5154718} = 0.5972188$ and $e^{-0.06082076} = 0.9409919$ times as before. i.e. will have approximately 6% to 40% of decrease.

(D)

Predict the probability of thermal distress at 31oF; the temperature at the time of the Challenger fligh.

Analysis

$$logit(\pi(TD = 1|Temperature)) = 15.0429 - 0.2322T \times 31 = 7.8447$$

$$\pi(TD = 1) = e^{7.8447} = 2552.172$$

$$\frac{p(TD = 1)}{1 - P(TD = 1)} = 2552.172$$

$$p(TD = 1) = 0.9996083$$

(E)

At what temperature does the predicted probability equal 0.5?

Analysis

$$odd(TD = 1) = \frac{p(TD = 1)}{1 - P(TD = 1)} = 1$$

$$logit(\pi(TD = 1|Temperature)) = log(1) = 0$$

$$15.0429 - 0.2322T \times Temperature = 0$$

$$Temperature = 64.78424$$

problem Two

(a)

Estimate β_1 and β_2 and interpret your result.

```
fit2 <- glm(cbind(adolesent$Yes,adolesent$No)~adolesent$Gender+adolesent$Race,family = binomial("logit"))
summary(fit2)
```

```
##
## Call:
## glm(formula = cbind(adolesent$Yes, adolesent$No) ~ adolesent$Gender +
##     adolesent$Race, family = binomial("logit"))
##
## Deviance Residuals:
##        1         2         3         4
## -0.08867   0.10840   0.14143  -0.13687
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.4555     0.2221  -2.050  0.04032 *
## adolesent$GenderMale   0.6478     0.2250   2.879  0.00399 **
## adolesent$RaceWhite   -1.3135     0.2378  -5.524 3.32e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 37.516984  on 3  degrees of freedom
## Residual deviance:  0.058349  on 1  degrees of freedom
## AIC: 25.186
##
## Number of Fisher Scoring iterations: 3
```

```
exp(coef(fit2))
```

```
##          (Intercept) adolesent$GenderMale  adolesent$RaceWhite
##            0.6341367            1.9113913            0.2688881
```

Analysis

1. the estimated value of $\beta_1 = 0.6478$, thie means if we fix race (i.e. let the race to be both black or both white), on average, for thoes who is male, their odd of having sexual intercourse would be $e^{0.6478} = 1.911331$ times the odd of having sexual intercourse for the famle.
2. the estimated value of $\beta_2 = -1.3135$, this means if we fix the gender factor (i.e. let the gender to be both male or both female), o2n average, for thoes who is white, their odd of having sexual intercourse would be $e^{-1.3135} = 0.2688773$ times the odd of having sexual intercourse for the black.

(b)

Construct a 95% confidence interval to describe the effect of gender on the odds of Intercourse controlling for race, Interpret your result

```
confint(fit2)
```

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

```
##                              2.5 %      97.5 %
## (Intercept)          -0.8971266 -0.02385449
## adolesent$GenderMale  0.2105773  1.09436472
## adolesent$RaceWhite  -1.7824267 -0.84865350
```

```
exp(confint(fit2))
```

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

```
##                          2.5 %    97.5 %
## (Intercept)          0.4077396 0.9764278
## adolesent$GenderMale 1.2343904 2.9872843
## adolesent$RaceWhite  0.1682294 0.4279908
```

Analysis

As we can see from the table, the confidence interval of β_1 is between 0.2105773 and 1.09436472. This indecate that, if we fix the gender factor (i.e. let the gender to be both male or both female) , on average we have 95 percent of confidenct to conclude that for thoes who is white, their odd of having sexual intercourse would be $e^{0.2105773} = 1.2343904$ to $e^{1.09436472} = 2.9872843$ times the odd of having sexual intercourse for the black.

(c)

Construct a 95% confidence interval to describe the effect of race on the odds of Intercourse controlling for gender, Interpret your result.

Analysis

As we can see from the table, the confidence interval of β_2 is between -1.7824267 and -0.84865350. This indecate that, if we fix race (i.e. let the race to be both black or both white), on average, we have 95 percent of confidenct to conclude that for thoes who is male, their odd of having sexual intercourse would be $e^{-1.7824267} = 0.1682294$ to $e^{-0.84865350} = 0.4279908$ times the odd of having sexual intercourse for the famle.

(d)

Test H0 : $\beta_1 = \beta_2 = 0$.

Analysis

Here we need to use the method of likelihood ratio test. Null deviance is Null deviance: 37.516984 on 3 degrees of freedom and Residual deviance: 0.058349 on 1 degrees of freedomn. Thus the test statistics is 37.516984 - 0.058349 = 37.45864. P = 2. We reject H0 since $37.45864 > \chi^2_2(0.95) = 5.99$. i.e. for race and gender, atl least one of them will have a significant inference on the whether having sexual intercourse

(e)

Test H0 : $\beta_1 = 0$

Analysis

As we can see from the table that the p-value of $\beta_1 = 0.00399 \leq 0.05$. And the z value is 2.879. Thus, we can reject the H0. i.e. for race will have a significant inference on the whether having sexual intercourse