Advanced Data Analysis HW5

Ao Liu, al3472

1.

(a)

Answer:

```
data = read.csv("Shuttle.csv", header = TRUE)
glm(ThermalDistress~Temperature, data = data, family = binomial("logit"))
```

```
  Call:  glm(formula = ThermalDistress ~ Temperature, family = binomial("logit"),
      data = data)

  Coefficients:
  (Intercept)  Temperature
      15.0429      -0.2322

  Degrees of Freedom: 22 Total (i.e. Null);  21 Residual
  Null Deviance:      28.27
  Residual Deviance: 20.32  AIC: 24.32
```

(b)

Answer:

According to the result that we got in (a), our estimation of β_1, the effect of temperature on the probability of thermal distress is

$$-0.2322$$

This implies that when we increase the temperature by 1 degree, the odds of having Thermal Distress changes by a multiplicative factor of $e^{-0.2322}$

(c)

Answer:

```
confint(glm(ThermalDistress~Temperature, data = data, family = binomial("logit")))
```

```
                2.5\%        97.5\%
(Intercept) 3.3305848    34.34215133
Temperature  -0.5154718   -0.06082076
```

According to the results in R, the 95% confidence interval for β_1 is

$$(-0.515718, -0.06082076)$$

so the the 95% confidence interval for e^{β_1} is

$$(0.597071743167396, 0.940991888047314)$$

This indicates that we are 95% confident that when we increase the temperature by 1 degree, the odds of having Thermal Distress changes by a multiplicative factor between 0.597071743167396 and 0.940991888047314.

(d)

Answer:

According to the estimation of the parameters, we have the following prediction function:

$$\hat{\pi}(TD|Temperature) = \frac{e^{15.0429-0.2322Temperature}}{1+e^{15.0429-0.2322Temperature}}$$

So when Temperature is 31 degree, we use the function above and get a prediction of the probability of Thermal Distress: 0.999608330327805.

(e)

Answer:

If the predicted probability equals to 0.5, then according to the function in (d), we have

$$2e^{15.0429-0.2322Temperature} = 1+e^{15.0429-0.2322Temperature}$$

After solving this equation, the Temperature is 64.7842377260982.

2.

(a)

Answer:

```
data = read.csv("/Users/lleiou/Google Drive/      Courses/4th term/Advanced-Data-Analysis/HW/Question/HW5/adolescent.csv")
logit = glm(cbind(Yes,No)~factor(Race)+factor(Gender), data = data, family = binomial)
logit
```

```
Call:  glm(formula = cbind(Yes, No) ~ factor(Race) + factor(Gender),
    family = binomial, data = data)

Coefficients:
      (Intercept)    factor(Race)White   factor(Gender)Male
          -0.4555              -1.3135               0.6478

Degrees of Freedom: 3 Total (i.e. Null);  1 Residual
Null Deviance:      37.52
Residual Deviance: 0.05835   AIC: 25.19
```

The estimation for β_1 is -1.3135, the estimation for β_2 is 0.6478.
If we hold the gender fixed, then we estimate the odds that 15 or 16 year-old white adolescents have sexual intercourse is $e^{-1.3135}$ times the odds that 15 or 16 year-old black adolescents have sexual intercourse. If we hold the race fixed, then we estimate the odds that 15 or 16 year-old male adolescents have sexual intercourse is $e^{0.6478}$ times the odds that 15 or 16 year-old female adolescents have sexual intercourse.

(b)

Answer:

```
confint  = confint(glm(cbind(Yes,No)~factor(Race)+factor(Gender), data = data, family = binomial))
exp(confint)
```

```
2.5 % 97.5 %
(Intercept) 0.4077396 0.9764278
factor(Race)White 0.1682294 0.4279908
factor(Gender)Male  1.2343904 2.9872843
```

The 95% confidence interval to describe the effect of gender on the odds of Intercourse controlling for race is:

$$(1.2343904, 2.9872843)$$

We are 95% confident that, if we hold the gender fixed, then we estimate the odds that 15 or 16 year-old white adolescents have sexual intercourse is between 1.2343904 and 2.9872843 times the odds that 15 or 16 year-old black adolescents have sexual intercourse.

(c)

Answer:

The 95% confidence interval to describe the effect of gender on the odds of Intercourse controlling for race is:

$$(0.1682294, 0.4279908)$$

We are 95% confident that if we hold the race fixed, then we estimate the odds that 15 or 16 year-old male adolescents have sexual intercourse is between 0.1682294 and 0.4279908 times the odds that 15 or 16 year-old female adolescents have sexual intercourse.

(d)

Answer:

```
summary(logit)
```

```
 Call:
glm(formula = cbind(Yes, No) ~ factor(Race) + factor(Gender),
    family = binomial, data = data)

 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 *
 factor(Race)White    -1.3135     0.2378  -5.524 3.32e-08 ***
 factor(Gender)Male    0.6478     0.2250   2.879  0.00399 **
---
 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
```

The test statistics = 37.516984 - 0.058349 = 37.458635. Since p=2 we rehject H_0 since $6.9674 > \chi^2_2(0.05) = 5.99$.

(e)

Answer:

From the result of (d), we see that the p-value for testing that $H_0 : \beta_1 = 0$ against $H_a : \beta_1 = 0$ is 3.32e-08. Therefore, we reject H_0.