Linear and Logistic Regression Questions version 2

Jared Cross

2024-05-05

Spaceship Titanic

Spaceship Titanic Home Planet

```
table(train$HomePlanet)
```

```
##
##          Earth Europa   Mars
##    201   4602   2131   1759
```

In the training set, there are 201 people whose home planet is not known, 4602 Earthlings, 2131 Europans and 1759 Martians.

Linear Regression

```
m1 = lm(Transported ~ HomePlanet,
        data=train)
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = Transported ~ HomePlanet, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6589 -0.4239  0.3412  0.4770  0.5760
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.51244    0.03461  14.807  < 2e-16 ***
## HomePlanetEarth  -0.08849    0.03536  -2.503   0.0123 *
## HomePlanetEuropa  0.14641    0.03620   4.044  5.3e-05 ***
## HomePlanetMars    0.01059    0.03653   0.290   0.7720
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4907 on 8689 degrees of freedom
## Multiple R-squared:  0.03738,    Adjusted R-squared:  0.03705
## F-statistic: 112.5 on 3 and 8689 DF,  p-value: < 2.2e-16
```

1

. The regression above uses an unknown home planet as a default. Please use it to predict the probability of being transported for Earthlings, Europans, and Martians.

2. Can you reject the null hypothesis that Earthlings are just as likely to be transported as people with an unknown home planet?

3. Can you reject the null hypothesis that Martians are just as likely to be transported as people with an unknown home planet?

Spa Treatments?

```
# Note:
# Spa_log = log(Spa + 1)

m2 = lm(Transported ~ Spa_log,
        data=train)

coef(m2)
```

```
## (Intercept)     Spa_log
##  0.62773608 -0.06600095
```

4. Please write an equation describing the relationship between amount spent at the spa and the chance of being transported. Note that "Spa_log" is $log_e(Spa + 1)$.

5. Predict the chance of being transported for someone who spent 0 on spa treatments.

6. Predict the chance of being transported for someone who spent 1000 on spa treatments.

```
m3 = lm(Transported ~
          Spa_log + CryoSleep,
        data=train)

coef(m3)
```

```
##   (Intercept)      Spa_log CryoSleepTRUE
##     0.42174220  -0.03145729    0.39555255
```

7. Using the multiple regression above, predict the chance of being transported for someone who spent 0 on spa treatments and who *was* in CryoSleep.

8. Predict the chance of being transported for someone who spent 1000 on spa treatments and who was *not* in CryoSleep.

Logistic Regression

```
m4 = glm(Transported ~ HomePlanet,
         data=train,
         family="binomial")

coef(m4)
```

```
##    (Intercept) HomePlanetEarth HomePlanetEuropa   HomePlanetMars
##     0.04976151     -0.35635638      0.60839260       0.04240145
```

9. Write an equation to predict the *log odds* of being transported based on HomePlanet (or if you prefer you can write one equation for each planet).

```
exp(coef(m4))
```

```
##    (Intercept) HomePlanetEarth HomePlanetEuropa   HomePlanetMars
##       1.051020        0.700223        1.837475         1.043313
```

10. Based on this logistic regression, write an equation to predict the odds of being transported based on HomePlanet.

11. Based on this logistic regression, predict the odds of being transported for an Earthling.

12. Based on this logistic regression, predict the probability of being transported for a Martian.

Bonus

```r
m5 = glm(Transported ~ HomePlanet +
           Spa_log,
         data=train,
         family="binomial")

coef(m5)
```

```
##     (Intercept)  HomePlanetEarth HomePlanetEuropa    HomePlanetMars
##       0.6077749       -0.4065392        1.3397906        -0.0962381
##         Spa_log
##      -0.3799279
```

```r
exp(coef(m5))
```

```
##     (Intercept)  HomePlanetEarth HomePlanetEuropa    HomePlanetMars
##       1.8363409        0.6659510        3.8182437         0.9082477
##         Spa_log
##       0.6839107
```

13. Based on the logistic regression above, predict the probabilty of being transported for an Earthling who spent 500 at the spa.