#8 Intro to Logistic Regression

Troy Edwards

2023-03-15

1

Import data.

```
donner <- read.csv("./donner.csv")
```

2

Fit a linear regression model using age to predict survival probability.

```
regres1 <- lm(survived ~ age, data = donner)
summary(regres1)
```

```
##
## Call:
## lm(formula = survived ~ age, data = donner)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.66886 -0.49521 -0.06775  0.45136  0.74524
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.869232   0.197106   4.410  6.8e-05 ***
## age         -0.013358   0.005777  -2.312   0.0256 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4794 on 43 degrees of freedom
## Multiple R-squared:  0.1106, Adjusted R-squared:  0.08992
## F-statistic: 5.347 on 1 and 43 DF,  p-value: 0.0256
```
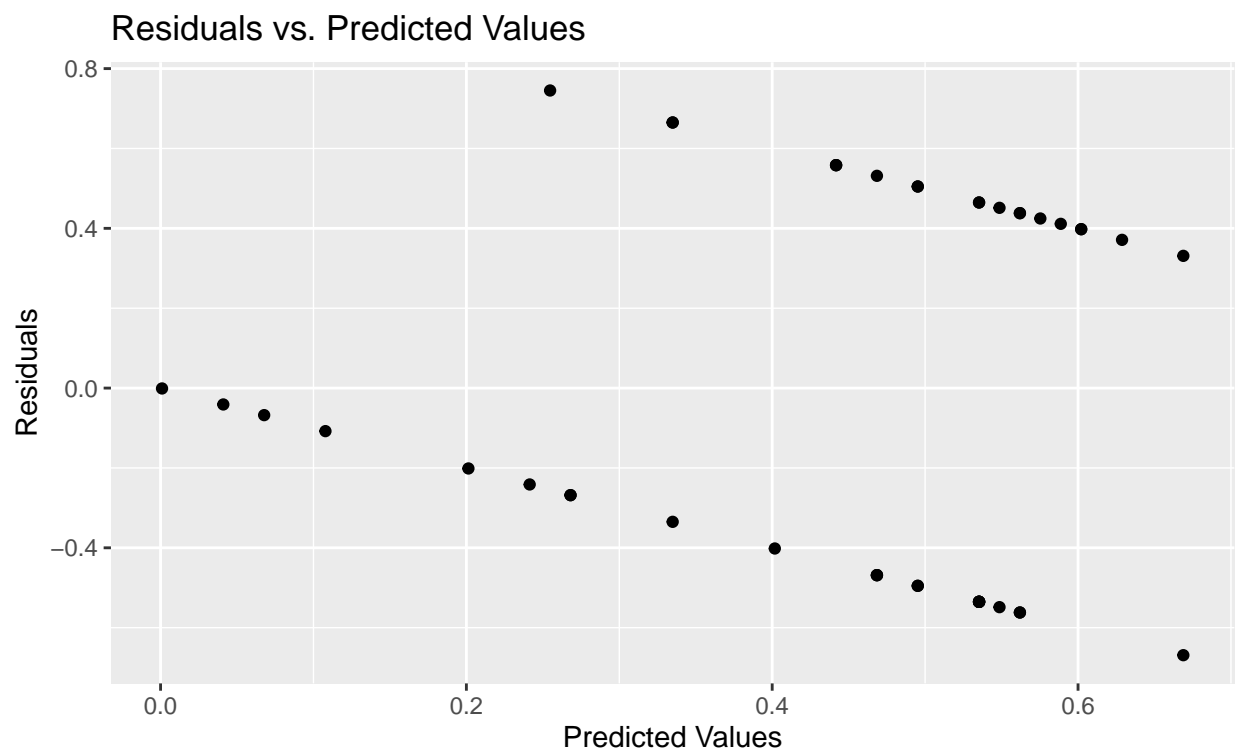
3

State regression equation.

$$\widehat{survived} = 0.869232 - 0.013358(age)$$

4

Check conditions/create a residual plot.

```
ggplot(donner, aes(predict(regres1), resid(regres1))) +
  geom_point() +
  labs(
    x = "Predicted Values",
    y = "Residuals",
    title = "Residuals vs. Predicted Values"
  )
```
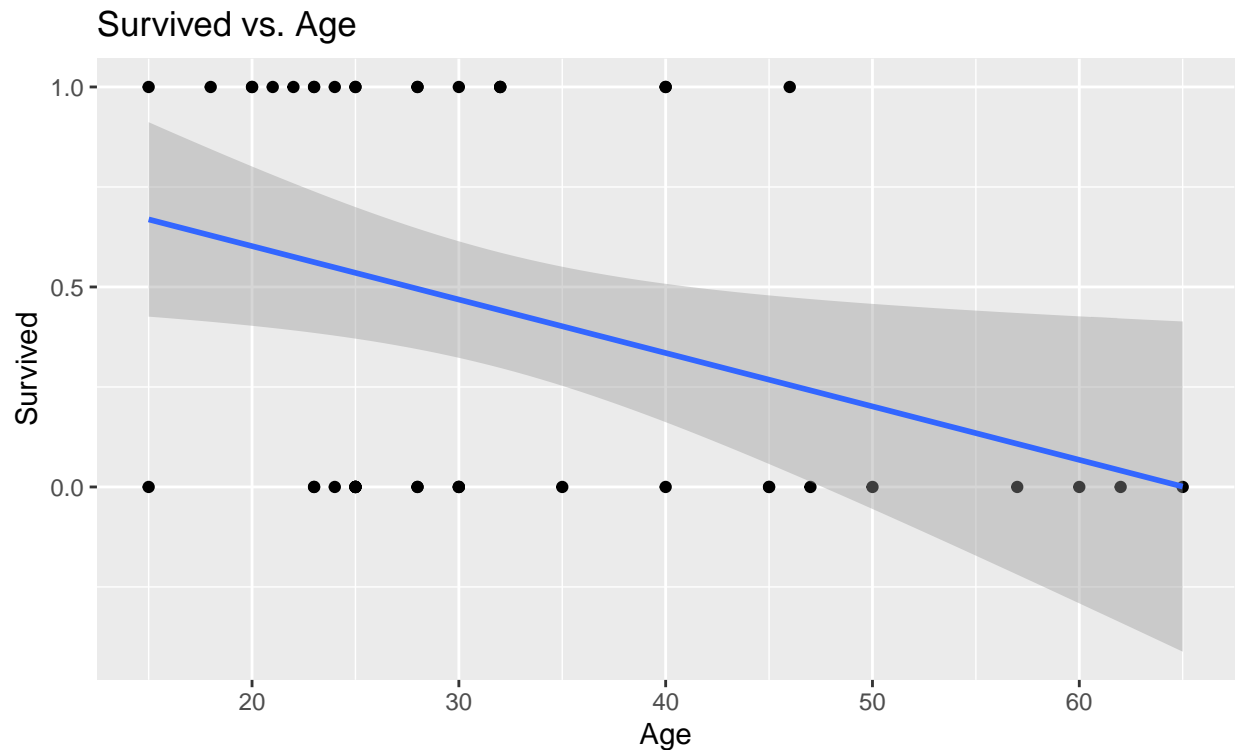

5

Construct a scatter plot to model your equation with the data included.

```
ggplot(donner, aes(age, survived)) +
  geom_point() +
```

```
geom_smooth(method = "lm") +
labs(
  x = "Age",
  y = "Survived",
  title = "Survived vs. Age"
)
```


6

Predict the survival probability of a 72 year old.

```
predict(regres1, newdata = data.frame(age = 72))
```

```
##           1
## -0.09255172
```

7

Interpret.

Since the predicted probability (-0.0926) is below 0, we can conclude that the probability of a 72 year old surviving is 0. I think that for the model to make sense, predicted values have to be clamped to the interval $[0, 1]$ such that any value greater than 1 would be interpreted as 1 and any value less than 0 would be interpreted as 0.