

Peer-graded Assignment: Predict the risk of getting esophageal cancer

Yang

11/10/2021

Overview

This is a supporting document of how to use the “assignment” shiny application. Specifically, this app is used to predict the risk of getting esophageal cancer, and the linear regression prediction model is generated using the data from a case-control study of esophageal cancer in Ile-et-Vilaine, France.

Links and the source code

The “assignment” shiny application can be found in <https://thefw21.shinyapps.io/Assignment/>, and the “ui.R” and “server.R” codes can be found in <https://github.com/ccm111222/Course-Project-Shiny-Application-and-Reproducible-Pitch>.

How to use the app

1. The “assignment” shiny application needs 3 inputs: age, the average daily alcohol consumption (gram), and the average daily tobacco consumption (gram). You can enter these information from the left panel of the app.
2. A multivariate linear regression will be run using the French esophageal cancer data, and a predicted risk of esophageal cancer (in proportion) will displayed, as well as the 95% confidence interval (CI).
3. The original French esophageal cancer data (small dots), as well as the data point that you just entered (a big dot), will appear as a 3D scatter plot at the bottom right. The predicted risk of getting esophageal cancer is illustrated using the color of the dot.

Appendix

The “ui.R” code

```
library(plotly)
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##   last_plot

## The following object is masked from 'package:stats':
##
##   filter

## The following object is masked from 'package:graphics':
##
##   layout
```

```
library(shiny)
shinyUI(fluidPage(
  titlePanel("Predicting the risk of getting esophageal cancer"),
  sidebarLayout(
    sidebarPanel(
      numericInput("age", "Please select your age:", value = 35, min = 25, max = 85, step = 1),
      numericInput("alcohol", "Please select your average daily alcohol consumption (gram):",
        value = 40, min = 0, max = 160, step = 1),
      numericInput("tobacco", "Please select your average daily tobacco consumption (gram):",
        value = 20, min = 0, max = 40, step = 1)
    ),
    # Show a plot of the generated distribution
    mainPanel(
      h3("The predicted risk of esophageal cancer"),
      textOutput("pred1"),
      h3("The predicted risk of esophageal cancer (the lower boundary of 95% CI)"),
      textOutput("pred2"),
      h3("The predicted risk of esophageal cancer (the upper boundary of 95% CI)"),
      textOutput("pred3"),
      h3("The 3D plot of the risk of esophageal cancer among the original data"),
      plotlyOutput("plot1")
    )
  )
))
```

The “server.R” code

```
library(shiny)
shinyServer(function(input, output) {
  esoph_processed<-esoph
  esoph_processed$risk<-(esoph_processed$ncases/(esoph_processed$ncases+esoph_processed$ncontrols))
  esoph_processed$age<-as.numeric(substr(esoph_processed$agegp,start = 1,stop = 2))
  esoph_processed$alcohol[esoph_processed$alcgp=="0-39g/day"]=0
  esoph_processed$alcohol[esoph_processed$alcgp=="40-79"]=40
  esoph_processed$alcohol[esoph_processed$alcgp=="80-119"]=80
  esoph_processed$alcohol[esoph_processed$alcgp=="120+"]=120
  esoph_processed$tobacco[esoph_processed$tobgp=="0-9g/day"]=0
  esoph_processed$tobacco[esoph_processed$tobgp=="10-19"]=10
```

```

esoph_processed$tobacco[esoph_processed$stobgp=="20-29"]=20
esoph_processed$tobacco[esoph_processed$stobgp=="30+"]=30
fit<-lm(risk~age+alcohol+tobacco, data=esoph_processed)

model_predict <- reactive({
  data1<-data.frame(age=input$age, alcohol=input$alcohol, tobacco=input$tobacco)
  predict(fit, newdata = data1, interval = "confidence")
})

output$pred1 <- renderText({
  model_predict()[1]
})

output$pred2 <- renderText({
  model_predict()[2]
})

output$pred3 <- renderText({
  model_predict()[3]
})

output$plot1 <- renderPlotly({
  library(plotly)
  data1<-data.frame(age=input$age, alcohol=input$alcohol, tobacco=input$tobacco)
  data2<-data.frame(age=input$age, alcohol=input$alcohol, tobacco=input$tobacco,
                    risk=predict(fit, newdata = data1))
  fig <- plot_ly(data=esoph_processed, x=~age, y=~alcohol, z=~tobacco,
                 type = "scatter3d", color = ~risk, name = "Original data", mode="markers")
  fig %>% add_trace(
    data=data2,
    x=~age,
    y=~alcohol,
    z=~tobacco,
    type = "scatter3d",
    color = ~risk,
    marker=list(size=20),
    name="Your data",
    mode="markers"
  )
})
})

```