P8110: Applied Regression II Spring 2024

Homework #6 [15 points]

Due on March 25, 11:59**AM**

The “hwdata2.csv” data contain the results from a study examining factors that influence the decision of whether to apply to graduate school. A total of 400 college juniors were asked if they are unlikely, somewhat likely, or very likely to apply to graduate school. The variables included are:

pared = 1 - at least one parent has a graduate degree, 0 - otherwise

public = 1 - the undergraduate institution is public, 0 - private

apply = 1 - unlikely, 2 - somewhat likely, 3 - very likely

count = frequency of each response category

1. Fit an ordinal logistic regression (proportional odds) model to the data to assess the response to whether to apply to graduate school for students with and without parent(s) having graduate degree, and in public and private institutions.
   1. Write down the model [2 points]

Ordinal logistic model for applying:

k = 3

* 1. Test the proportional odds assumption. Show hypotheses, test statistic, degrees of freedom, p-value, and conclusion. [2 points]

Hypotheses:

**A screenshot of a computer

Description automatically generated**



**Conclusion: We fail to reject H0 at α = 0.05. The proportional odds assumption is reasonable.**

* 1. Estimate the odds ratio and 95% CI of a lower rating (i.e. rating less likely) regarding whether to apply to graduate school between students with and without parent(s) having graduate degree. Do students with parent(s) having graduate degree less or more likely to apply to graduate school? Justify your conclusion. [3 points]

**A close-up of numbers

Description automatically generated**



**The estimated odds ratio and 95% CI of a lower rating is 0.326[0.194, 0.547]. In other words, students with at least one parent having a graduate degree are 0.33 times more likely to apply to graduate school compared to students with no parents having a graduate degree, as the estimated odds is greater than one.**

* 1. Estimate the probability of rating “very likely” regarding whether to apply to graduate school for students with parent(s) having graduate degree, and in a private institution. [Hint: Pr(*y* = 3) = 1 − Pr(*y*≤ 2)] [2 points]

A screenshot of a computer

Description automatically generated



A screenshot of a computer code

Description automatically generated



1. Fit a multinomial logistic regression model to the data using “unlikely” as the reference category.
   1. Write down the model. [2 points]

k = 3

* 1. Estimate the odds ratio and 95% CI of rating “very likely” versus “unlikely” between students with and without parent(s) having graduate degree. [2 points]

A number and numbers on a white background

Description automatically generated



The estimated odds and 95% confidence interval of rating “very likely” versus “unlikely” between students with and without parent(s) having graduate degree is 4.582[2.036, 10.314].

P8110: Applied Regression II Spring 2024

* 1. Estimate the probability of rating “very likely” regarding whether to apply to graduate school for students with parent(s) having graduate degree, and in a private institution. [2 points] A screenshot of a computer

     Description automatically generated

A screenshot of a computer

Description automatically generated



**The probability of rating “very likely” regarding whether to apply to graduate school for students with parent(s) having a graduate degree and in a private institution is 17.8%.**

**R code:**

**---**

**title: "HW6\_Okonkwo\_co2554"**

**author: "Camille Okonkwo"**

**date: "2024-03-20"**

**output: github\_document**

**---**

**```{r setup, include=FALSE}**

**knitr::opts\_chunk$set(echo = TRUE)**

**library(tidyverse)**

**```**

**# Prepare the data**

**```{r}**

**hwdata2 = read.csv("data/hwdata2.csv")**

**hwdata2.1 = uncount(hwdata2, weights = count)**

**# re-factor variables**

**hwdata2.1$pared2 = relevel(as.factor(hwdata2.1$pared),**

**ref="0")**

**hwdata2.1$public2 = relevel(as.factor(hwdata2.1$public),**

**ref="0")**

**hwdata2.1$apply = as.factor(hwdata2.1$apply)**

**table(hwdata2.1$apply)**

**```**

**# Fit the ordinal logistic model**

**```{r}**

**library(VGAM)**

**# proportional odds model**

**fit = vglm(apply ~ pared2 + public2,**

**data = hwdata2.1,**

**family = cumulative("logitlink",**

**parallel = TRUE))**

**summary(fit)**

**```**

**# Test the proportional odds assumption**

**```{r}**

**# non- proportional odds model**

**fit2 = vglm(apply ~ pared2 + public2,**

**data = hwdata2.1,**

**family = cumulative("logitlink",**

**parallel = FALSE))**

**summary(fit2)**

**lrtest(fit2, fit)**

**```**

**# Calcualte odds ratio & 95% CI of a lower rating**

**```{r}**

**exp(cbind(OR = coef(fit), confint(fit)))**

**```**

**# Estimate the probability of rating “very likely”**

**```{r}**

**# predicted probabilities**

**newdata = data.frame(pared2=c("0","1"),**

**public2=c("1", "0"))**

**pred = predict(fit, newdata=newdata, "response")**

**cbind(newdata,pred)**

**```**

**# Fit a multinomial logistic regression model to the data using “unlikely” as the reference category.**

**```{r}**

**fit3 = vglm(apply ~ pared2 + public2,**

**data = hwdata2.1,**

**family = multinomial(refLevel = "1"))**

**summary(fit3)**

**# odds ratio and 95% CI**

**exp(cbind(OR=coef(fit3),confint(fit3)))**

**# predicted probabilities**

**d\_hwdata2.1 = data.frame(pared2=rep(c("0","1"),2),**

**public2=rep(c("1","0"),2))**

**pred2 = predict(fit3, newdata=d\_hwdata2.1, "response")**

**cbind(d\_hwdata2.1,pred2)**

**```**