HW4

Carina\_Sun

2024-04-11

knitr::opts\_chunk$set(echo = TRUE)  
  
library(MASS)  
library(psych)  
library(nnet)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ ggplot2::%+%() masks psych::%+%()  
## ✖ ggplot2::alpha() masks psych::alpha()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ dplyr::select() masks MASS::select()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tinytex)  
library(rmarkdown)  
library(skimr)  
library(reshape2)

##   
## Attaching package: 'reshape2'  
##   
## The following object is masked from 'package:tidyr':  
##   
## smiths

data <-read.csv("ORCouponApply.csv", stringsAsFactors = T)  
  
#1.Check the structure of the data using str  
str(data)

## 'data.frame': 400 obs. of 5 variables:  
## $ X : int 290 373 41 49 157 295 18 177 385 234 ...  
## $ rpurchase: Factor w/ 3 levels "Somewhat Likely",..: 2 2 2 2 2 2 2 2 1 2 ...  
## $ coupon : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ peers : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ quality : num 1.89 1.97 2.02 2.08 2.12 2.14 2.16 2.21 2.21 2.22 ...

#2. Change purchase probability to an ordered factor   
data$rpurchase <- factor(data$rpurchase, levels = c ("Unlikely","Somewhat Likely", "Very Likely"))  
  
#3. Recheck the structure of the data to make sure it is now ordered   
str(data)

## 'data.frame': 400 obs. of 5 variables:  
## $ X : int 290 373 41 49 157 295 18 177 385 234 ...  
## $ rpurchase: Factor w/ 3 levels "Unlikely","Somewhat Likely",..: 1 1 1 1 1 1 1 1 2 1 ...  
## $ coupon : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ peers : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ quality : num 1.89 1.97 2.02 2.08 2.12 2.14 2.16 2.21 2.21 2.22 ...

levels(data$rpurchase)

## [1] "Unlikely" "Somewhat Likely" "Very Likely"

#4. Describe the data using any descriptive technique you would like   
describe(data)

## vars n mean sd median trimmed mad min max range skew  
## X 1 400 200.50 115.61 200.50 200.50 148.26 1.00 400.00 399.0 0.00  
## rpurchase\* 2 400 1.55 0.67 1.00 1.44 0.00 1.00 3.00 2.0 0.82  
## coupon 3 400 0.16 0.36 0.00 0.07 0.00 0.00 1.00 1.0 1.87  
## peers 4 400 0.14 0.35 0.00 0.05 0.00 0.00 1.00 1.0 2.04  
## quality 5 400 2.99 0.40 2.98 2.99 0.40 1.89 3.99 2.1 -0.03  
## kurtosis se  
## X -1.21 5.78  
## rpurchase\* -0.48 0.03  
## coupon 1.51 0.02  
## peers 2.16 0.02  
## quality -0.43 0.02

#5. Using the ORCoupon.csv data set to determine if your IV’s coupon, peers, and quality has any effect on the probability of purchasing (DV)  
model <- polr(rpurchase~coupon+peers+quality, data=data, Hess= T)  
model

## Call:  
## polr(formula = rpurchase ~ coupon + peers + quality, data = data,   
## Hess = T)  
##   
## Coefficients:  
## coupon peers quality   
## 1.04770356 -0.05881433 0.61596620   
##   
## Intercepts:  
## Unlikely|Somewhat Likely Somewhat Likely|Very Likely   
## 2.197833 4.293281   
##   
## Residual Deviance: 717.0249   
## AIC: 727.0249

#It tells us proportional odds logistic regression. It tells us the odds ratio, the change in the log odds of being in a higher category as the predictor increases. It also tells the intercepts, the threshold for each comparison.  
  
#6.Exponentiate the coefficients   
exp(coef(model))

## coupon peers quality   
## 2.8510962 0.9428818 1.8514446

#Customers with a coupon are approximately 2.85 times more likely than those without coupon to be in a higher category of purchase probability (from 'Unlikely' to 'Somewhat Likely' or from 'Somewhat Likely' to 'Very Likely').  
#The odds of moving to a higher category of purchase probability decrease by an odds of about 0.94 for products that are peer-recommended, compared to products without peer recommendation  
#For every one-unit increase in the quality score of the product, the odds of moving to a higher category of purchase probability (from 'Unlikely' to 'Somewhat Likely' or from 'Somewhat Likely' to 'Very Likely') increase by an odd of 1.85.  
  
#7.  
summary(model)

## Call:  
## polr(formula = rpurchase ~ coupon + peers + quality, data = data,   
## Hess = T)  
##   
## Coefficients:  
## Value Std. Error t value  
## coupon 1.04770 0.2658 3.9419  
## peers -0.05881 0.2979 -0.1975  
## quality 0.61597 0.2606 2.3633  
##   
## Intercepts:  
## Value Std. Error t value  
## Unlikely|Somewhat Likely 2.1978 0.7770 2.8287  
## Somewhat Likely|Very Likely 4.2933 0.8018 5.3546  
##   
## Residual Deviance: 717.0249   
## AIC: 727.0249

fit <- deviance(multinom(rpurchase~1, data=data)) - deviance(model)

## # weights: 6 (2 variable)  
## initial value 439.444915   
## final value 370.602641   
## converged

pchisq(fit, 2, lower.tail=FALSE)

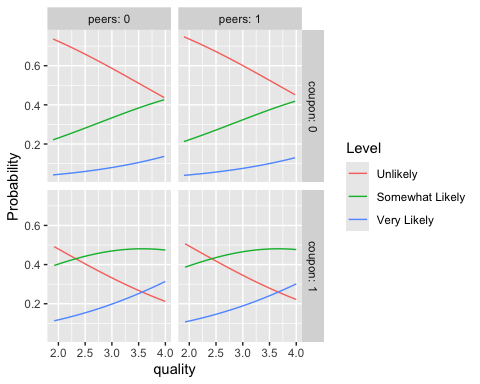
## [1] 5.614235e-06

confint(model)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## coupon 0.5281768 1.5721750  
## peers -0.6522060 0.5191385  
## quality 0.1076201 1.1309148

#8. Create a new dataframe with peers, coupon and quality and call it newdat and use the code below:  
newdat <- data.frame (coupon = rep(0:1, 200), peers = rep(0:1, each = 200), quality = rep(seq(from = 1.9, to = 4, length.out = 100), 4))  
  
#9. CBind the probabilities from the ordinal model called model to the newdat dataframe  
newdat <- cbind(newdat, predict(model, newdat, type = "probs"))  
  
#10. Use the reshape2 package to “melt” the data from the newdat dataframe for the levels of rpurchase (unlikely, somewhat likely, very unlikely) into one column and assign it to the object name lnewdat.  
lnewdat <- melt(newdat, id.vars = c("coupon", "peers", "quality"),  
 variable.name = "Level", value.name="Probability")  
  
#11. Create a ggplot using the aes as x = quality, y = Probability, and colour = Level from the variable.name from the lnewdat dataframe you created. aes(x = quality, y = Probability, colour = Level)) + geom\_line() + facet\_grid(coupon ~ peers, labeller="label\_both").   
ggplot(lnewdat, aes(x = quality, y = Probability, colour = Level)) +  
 geom\_line() +  
 facet\_grid(coupon ~ peers, labeller = "label\_both")



# Peers Recommendation is not a significant factor in predicting the repurchase behavior since the charts of with recommendations or not are very similar.   
# When having coupons, the repurchase behavior generally increases. As the quality improves, the repurchase behavior increases.