Lab 12: Multinomial Logistic Regression

SDS358: Applied Regression Analysis

Michael J. Mahometa, Ph.D.

.

> "Your job is not to dumb things down; it's to make everyone in the room smarter."
>
> *Edward Tufte*

Introduction

The basic idea of Lab is as follows: Answer a research question with the provided dataset. Each week, that research question (and data) will change depending on the topic we've covered the prior class days. Once we're done with Lab, you'll have a Lab Assignment, that will look a lot like the Lab: a research question you'll need to answer given some data. In Lab, you'll learn the procedure for answering the research question. For the Lab Assignment, you'll do that procedure for a grade (independently).

To help answer the research question, we'll follow some basic steps that we'll repeat throughout the semester:

- Reflect on the Question: Figure out the variables of interest, and the technique that's required.
- Analyze the Data: Perform the steps required for the technique.
- Draw Conclusions: Use the information that you got from the prior step to answer the research question in a concise, logical manner.

Let's get started:

Primary Research Question:

What item characteristic: Usefulness, Packaging, Price, or Gender of purchaser, best predicts the outcome of purchasing the item, placing the item in a wish list, or not buying an item? Explain and graph your findings.

Step1: Reflect on the Question:

Download the syntax and data files from Canvas.

Let's load in our SDSRegressionR package so that we can use some of it's functions later:

```
#Load our class package
library(SDSRegressionR)
```

Next, we'll load in the data. Be sure to use the basic file structure we talked about the first Lab: Put your syntax in a folder specific to this Lab. Then, make a "data" folder in that same place - use lowercase. If you do that, then all of this syntax will work like a charm.

```
buy <- read_csv("data/purchases.csv")
```

Check the Data:

To make sure that we're working with the right data, and that we're all looking at it the same way, we'll answer some basic questions about the data before moving on:

1. How many observations are in the dataset for the model?
2. What was the highest Usefulness rating in the first 10 observations?
3. Of the first 10 observations, how many resulted in a purchase?

These questions can be answered simply by looking at the dataset once it's loaded in:

```
View(buy)
```

Check the Variables of Interest

Let's find the variables that we need to answer the primary research question:

1. Which variable tells us the Outcome variable for this model?
   - What type of variable is this?
2. What are the variables of interest for the model?
3. Classify each of the variables of interest.

Again, these can be answered by looking at the dataframe, and with the help of the *names()* function. Also, the codebook for the data frame is our friend. You can open this in R or Excel. Remember, R is case-sensitive.

```
names(buy)
```

```
## [1] "PurchaseNumber" "Purchase"       "Usefulness"     "Packaging"
## [5] "Price"          "Gender"
```

Reflect on the Method

The last part of Reflect on the Question asks about the method or technique we'll use.

1. We will use (Ordinal or) Multinomial Logistic Regression to answer this question. Why?
2. For Multinomial, we'll need to run the model more than once. Why?
3. We'll use a graph of predicted probabilities. Why?

Step2: Analyze the Data

In this step, we'll run the provided syntax and answer some questions about the output to help us prepare for the final step.

Here's the syntax you'll need (from the .R syntax file):

```
#### Here is the R script you will use:  (remember that # indicates a comment) ####
#Lab12: Ordinal/Multinomial Logistic Regression

library(SDSRegressionR)

#Data
buy <- read_csv("data/purchases.csv")

#Examine
table(buy$Purchase) #Order is backwards
table(buy$Gender)

#Factor
buy <- buy %>%
   mutate(Purchase = factor(Purchase, levels=c("Leave Item", "Wish List", "Buy Item")),
          Gender = factor(Gender, levels=c("Male", "Female")))

#Model
library(nnet)
m_buy <- multinom(Purchase ~ Usefulness + Packaging + Price + Gender, data = buy)
summary(m_buy)

#Overall
x2 <- deviance(multinom(Purchase~1, data=buy)) - deviance(m_buy)
x2
pchisq(x2, 6, lower.tail=FALSE)

#Individual paramerters
z <- summary(m_buy)$coefficients/summary(m_buy)$standard.errors
z
p <- (1 - pnorm(abs(z), 0, 1))*2
round(p, 6)

#Odds-ratios
exp(coef(m_buy))

#Change the baseline:
buy <- buy %>%
   mutate(Purchase_Leave = factor(buy$Purchase,
                                  levels=c("Buy Item", "Leave Item", "Wish List")))

#Model_Leave
m_buy_l <- multinom(Purchase_Leave ~ Usefulness + Packaging +
                       Price + Gender, data = buy)
summary(m_buy_l)

#Individual paramerters_Leave
```

```
z <- summary(m_buy_l)$coefficients/summary(m_buy_l)$standard.errors
z
p <- (1 - pnorm(abs(z), 0, 1))*2
round(p, 6)

#Odds-ratios_Leave
exp(coef(m_buy_l))

#Graphing
library(emmeans)
use_mns <- summary(emmeans(m_buy, c("Purchase", "Usefulness"),
                           at=list(Usefulness=seq(0,10,1))))

ggplot(use_mns, aes(y=prob, x=Usefulness, color=Purchase)) +
  geom_line() +
  labs(title="Usefulness impact") +
  theme_bw()

price_mns <- summary(emmeans(m_buy, c("Purchase", "Price"),
                             at=list(Price=seq(0,10,1))))

ggplot(price_mns, aes(y=prob, x=Price, color=Purchase)) +
  geom_line() +
  labs(title="Price impact") +
  theme_bw()
```

Question 1

The overall model was significant in the prediction of Purchasing Outcome: LR chi2 (_____) = _____, $p < 0.05$.

Question 2

The baseline outcome for this model is: _____.

Question 3

As Usefulness increases by one whole unit, the odds of Leaving the Item over Buying the Item decreased by _____ %.

Question 4

As Usefulness increases by one whole unit, the odds of Buying the Item over Leaving the Item increased by _____ %.

Question 5

Of the three predictors, _____ was significantly related both Buying an Item over Leaving an Item (OR = _____, z = _____, p = _____), as well as putting an item on a Wish List over Leaving an Item (OR

= _____, z = _____, p = _____).

Question 6

With a Usefulness score of 7, the predicted probability of Buying an item is _____, while Leaving an Item is _____, and putting an item on a Wish List is _____.

Question 7

Even with a Usefulness score of 10, the predicted probability of Buying an item is _____, while Leaving an Item is _____, and putting an item on a Wish List is _____.

Question 8

What "minimum" Usefulness score would you need your item to be above to drive purchasing the item?

Step3: Draw Conclusions

The final step is for us to Draw Conclusions. We'll take the syntax we've been given from Analyze the Question, run it, then examine the output. The questions from the prior step help set us up for the Draw Conclusions part.

We'll "fill in the blanks" in a canned paragraph for the Lab. For the Lab Assignment, you'll need to come up with a similar paragraph all on your own (please don't steal mine).

Primary Research Question
What item characteristic: Usefulness, Packaging, or Price, best predicts the outcome of purchasing the item, placing the item in a wish list, or not buying an item? Explain and graph your findings.

> Multinomial Logistic Regression analysis was used to examine the product perceived Usefulness, Packaging design, and Price on Purchasing outcome (Buy, Leave, or Wish List an item). The overall model for training data was significant overall (LR chi2(3) = _____, $p < 0.05$). Of the three predictors in the model, only Useful of the item was significant for an effect between Leave and Buy (OR = _____, z = _____, $p < 0.05$), as well as Leave and Wish List (OR = _____, z = _____, $p < 0.05$). However, although not significant for both transitions, Price of the item showed an effect for Leave and Wish List (OR = _____, z = _____, $p < 0.05$), and a marginal effect for Leave and Buy (OR = _____, z = _____, p = _____). See graphs for predicted probabilities for each outcome for both Usefulness and Price.

Lab Assignment

Now, with the tools at your disposal (the R syntax from Lab, and the logic of proceeding through the three steps of answering the research question), you'll have a Lab Assignment to complete (independently). For now, the Lab Assignment is to be completed in Canvas. It will follow the basic structure, and lead to the same place - answering the research question with a concise paragraph as in Draw Conclusions.

Good Luck!