ASSIGNMENT 6 - Penthalon

Section 81 Dain Hall Trevor Cornell

Preliminaries

Determine notebook defaults:

Load packages:

Read in the data:

```
# use load("filename.Rdata") for .Rdata files
load("/Users/dain/Programs/R_Projects/MKTG_482_HW6/PentathlonTargeting.RData")

pent.nptb <- pent %>%
   mutate(buyer=factor(buyer), female=factor(female), age=factor(age), message=factor(message))
```

Assignment answers

Part 1

Question 1

For each customer, determine the action (a message for endurance, strength, water, team, backcountry, racquet, or the no-message control condition) that is predicted to lead to the highest probability of purchase. Describe what approach you took to predict probability of purchase.

```
pent.nptb.train <- pent.nptb %>% filter(training==1)
pent.nptb.test <- pent.nptb %>% filter(training==0)

lr.formula <- formula(buyer ~ age + female + income + education + children + freq_endurance + freq_stren

lr.team <- glm(lr.formula, family=binomial, data=pent.nptb.train %>% filter(message=="team"))
lr.backcountry <- glm(lr.formula, family=binomial, data=pent.nptb.train %>% filter(message=="backcountry
lr.endurance <- glm(lr.formula, family=binomial, data=pent.nptb.train %>% filter(message=="endurance"))
lr.water <- glm(lr.formula, family=binomial, data=pent.nptb.train %>% filter(message=="water"))
lr.racquet <- glm(lr.formula, family=binomial, data=pent.nptb.train %>% filter(message=="racquet"))
lr.strength <- glm(lr.formula, family=binomial, data=pent.nptb.train %>% filter(message=="strength"))
lr.control <- glm(lr.formula, family=binomial, data=pent.nptb.train %>% filter(message=="control"))

pent.nptb.test <- pent.nptb.test %>%
   mutate(
      pr.team = predict(lr.team, newdata=pent.nptb.test, type="response"),
      pr.backcountry = predict(lr.backcountry, newdata=pent.nptb.test, type="response"),
      pr.endurance = predict(lr.endurance, newdata=pent.nptb.test, type="response"),
      pr.water = predict(lr.water, newdata=pent.nptb.test, type="response"),
      pr.racquet = predict(lr.racquet, newdata=pent.nptb.test, type="response"),
      pr.strength = predict(lr.strength, newdata=pent.nptb.test, type="response"),
```

```
      pr.control = predict(lr.control, newdata=pent.nptb.test, type="response"),
   )

pent.nptb.test <- pent.nptb.test %>%
   mutate(
      pr.max = pmax(pr.team, pr.backcountry, pr.endurance, pr.water, pr.racquet, pr.strength, pr.control
      pr.offer = case_when(
         pr.team == pr.max ~ "team",
         pr.backcountry == pr.max ~ "backcountry",
         pr.endurance == pr.max ~ "endurance",
         pr.water == pr.max ~ "water",
         pr.racquet == pr.max ~ "racquet",
         pr.strength == pr.max ~ "strength",
         pr.control == pr.max ~ "control",
      )
   )
pent.nptb.test %>% head(5) %>% select(custid, pr.max, pr.offer)
```

```
# A tibble: 5 x 3
  custid  pr.max pr.offer
   <dbl>   <dbl> <chr>
1      3 0.00601 endurance
2     25 0.0197  endurance
3     30 0.0580  strength
4     55 0.00624 strength
5     97 0.0247  endurance
```

Question 2

Report for each action the percent of customers in the test sample for whom that action maximizes their probability of purchase.

```
pent.nptb.test %>% tabyl(pr.offer)
```

```
    pr.offer      n     percent
 backcountry   1447 0.008038889
   endurance 125861 0.699227778
     racquet  12299 0.068327778
    strength  36584 0.203244444
        team   1680 0.009333333
       water   2129 0.011827778
```

Question 3

For each customer, determine the action (a message for endurance, strength, water, team, backcountry, racquet, or the no-message control condition) that is predicted to lead to the highest predicted profit (the COGS is 60%). Heads-up: There are different ways to predict order size; pick one that you think predicts order size the best. Explain how you calculated expected profit.

```
osavg.team <- mean(filter(pent.nptb.test, total_os!=0 & message == "team")$total_os)
osavg.backcountry <- mean(filter(pent.nptb.test, total_os!=0 & message == "backcountry")$total_os)
osavg.endurance <- mean(filter(pent.nptb.test, total_os!=0 & message == "endurance")$total_os)
osavg.water <- mean(filter(pent.nptb.test, total_os!=0 & message == "water")$total_os)
osavg.racquet <- mean(filter(pent.nptb.test, total_os!=0 & message == "racquet")$total_os)
osavg.strength <- mean(filter(pent.nptb.test, total_os!=0 & message == "strength")$total_os)
```

```
osavg.control <- mean(filter(pent.nptb.test, total_os!=0 & message == "control")$total_os)

margin <- 0.4
pent.nptb.test <- pent.nptb.test %>%
   mutate(
      ep.team=pr.team * osavg.team * margin,
      ep.backcountry=pr.backcountry * osavg.backcountry * margin,
      ep.endurance=pr.endurance * osavg.endurance * margin,
      ep.water=pr.water * osavg.water * margin,
      ep.racquet=pr.racquet * osavg.racquet * margin,
      ep.strength=pr.strength * osavg.strength * margin,
      ep.control=pr.control * osavg.control * margin,
   ) %>%
   mutate(
      ep.max = pmax(ep.team, ep.backcountry, ep.endurance, ep.water, ep.racquet, ep.strength, ep.contro
      ep.offer = case_when(
         ep.team == ep.max ~ "team",
         ep.backcountry == ep.max ~ "backcountry",
         ep.endurance == ep.max ~ "endurance",
         ep.water == ep.max ~ "water",
         ep.racquet == ep.max ~ "racquet",
         ep.strength == ep.max ~ "strength",
         ep.control == ep.max ~ "control",
      )
   )
pent.nptb.test %>% head(5) %>% select(custid, ep.max, ep.offer)
```

```
# A tibble: 5 x 3
  custid ep.max ep.offer
   <dbl>  <dbl> <chr>
1      3  0.143 racquet
2     25  0.423 endurance
3     30  1.26  strength
4     55  0.136 strength
5     97  0.587 water
```

Profit Formula = Predicted purchase probability * Avg Order Size for bucket * Margin *Predicted purchase probability:* using logistc regression and for each message *Avg Order size:* calculated by bucketing customers according to prior purchases. *Margin:* 1 - COGS %

Question 4

Report for action the percent of customers in the test sample for whom that action maximizes their predicted profit.

```
pent.nptb.test %>% tabyl(ep.offer)
```

```
    ep.offer     n    percent
 backcountry  9686 0.05381111
   endurance 72258 0.40143333
     racquet 58646 0.32581111
    strength 22030 0.12238889
        team  5564 0.03091111
       water 11816 0.06564444
```

Question 5

Using the predicted profit for all consumers in the test sample, what profit can we obtain on average per customer when we customize the message to each customer (including potentially sending no message)?

```
mean(pent.nptb.test$ep.max)
```

```
[1] 0.6729486
```

Question 6

Using the predicted profit for all consumers in the test sample, what profit can Pentathlon obtain on average per customer if every customer receives the same message (or the no-message control condition)? Answer the question for each of the seven possible actions (a message for endurance, strength, water, team, backcountry, racquet, or the no-message control condition).

```
pent.nptb.test %>%
   summarise_at(vars(ep.endurance, ep.strength, ep.water, ep.team, ep.backcountry, ep.racquet, ep.contr
```

```
# A tibble: 1 x 7
  ep.endurance ep.strength ep.water ep.team ep.backcountry ep.racquet ep.control
         <dbl>       <dbl>    <dbl>   <dbl>          <dbl>      <dbl>      <dbl>
1        0.598       0.575    0.617   0.575          0.549      0.566      0.429
```

Question 7

Using the predicted profit for all consumers in the test sample, what profit can Pentathlon obtain on average per customer if every customer is assigned at random to receive one of the six messages?

```
pent.nptb.test %>%
   summarise_at(vars(ep.endurance, ep.strength, ep.water, ep.team, ep.backcountry, ep.racquet, ep.contr
   rowMeans()
```

```
[1] 0.5584633
```

Question 8

Based on the numbers calculated in question 5 and 6, for the typical promotional e-mail blast to 5,000,000 customer, what improvement (in percent and in total Euro) does Pentathlon expect to get from customizing the message to each customer rather than assigning customers the message that is most profitable on average?

```
inc_profit <- 5000000 * mean(pent.nptb.test$ep.max) - 5000000 * mean(pent.nptb.test$ep.water)
inc_percentage <- mean(pent.nptb.test$ep.max) / mean(pent.nptb.test$ep.water) - 1

paste("Incremental Profit of ep.max v ep.water in Euros: ", dollar(inc_profit),
      "Incremental % of ep.ax v ep.water: ", percent(inc_percentage, 0.1))
```

```
[1] "Incremental Profit of ep.max v ep.water in Euros:  $279,037 Incremental % of ep.ax v ep.water:  9.
```

Part 2

Comment on the draft for a new e-mail policy proposal. Are there any weaknesses? Can you suggest at least one improvement?

> The proposed email policy is good. However, if profit margins remain constant, some customers will never receive different emails, missing opportunities for sales and creating poor data for future campaigns. Some amount of randomization and control groups should be maintained. This is a result of rerunning the same analysis monthly as opposed to randomized experiment. Therefore, you cannot attribute causation because the email assignments are no longer random.

Email:

A. Promotional e-mails will be allocated to departments on a monthly basis. B. For the first month after this policy goes into effect we assign customer e-mails to departments as follows: a. For each customer, the analytics team forecasts the actions that yield the highest and the second highest expected profit among the six possible messages and not sending a message. b. The two departments (or no department if the control was among the two best choices) who yield the highest and the second highest expected profit for a customer each control 1/2 of the allowed e-mail messages to that customer during that month. C. During the last week of each subsequent month the analytics team uses the data from e-mails sent during the first three weeks in that month and > repeats the analysis described in step 2.