

# Appendix-B

November 4, 2024

```
[20]: import os
import numpy as np
import pandas as pd
import math
import matplotlib.pyplot as plt
import seaborn as sns
from IPython.display import display
from sklearn.linear_model import LogisticRegression
```

```
[21]: df = pd.read_csv("Simmons-data-raw.csv")
df.head(5)
```

```
[21]:
```

	Customer	Spending(000)	Card	Coupon-Usage-Indicator
0	1	2.291	1	0
1	2	3.215	1	0
2	3	2.135	1	0
3	4	3.924	0	0
4	5	2.528	1	0

```
[22]: df.dtypes
```

```
[22]: Customer                int64
Spending(000)              float64
Card                      int64
Coupon-Usage-Indicator    int64
dtype: object
```

## 0.1 B-1

```
[24]: predictor_cols = ["Spending(000)", "Card"]
target_col = "Coupon-Usage-Indicator"

model = LogisticRegression()
model.fit(df[predictor_cols].values, df[target_col])
beta0 = model.intercept_[0]
beta1 = model.coef_[0][0]
beta2 = model.coef_[0][1]
print('LR coefficients:')
```

```

print('BETA0 (or constant term): {:.4f}'.format(beta0))
print('BETA1 (coeff. For X1): {:.4f}'.format(beta1))
print('BETA2 (coeff. For X2): {:.4f}'.format(beta2))
print('\n')
print('ODDS RATIOS:')
print('X1: {:.4f}'.format(np.exp(beta1)))
print('X2: {:.4f}'.format(np.exp(beta2)))

```

LR coefficients:

BETA0 (or constant term): -2.0067

BETA1 (coeff. For X1): 0.3299

BETA2 (coeff. For X2): 0.9179

ODDS RATIOS:

X1: 1.3908

X2: 2.5040

## 0.2 B-2

```

[26]: jack=[ [2.0, 1.0] ]
      jill=[ [4.0, 0.0] ]
      def predict_coupon_usage(X):
          pred_val = model.predict_proba(X)[: ,1]
          return(pred_val)

      print("Probability of Response from Jack = {:.4f}".
            ↪format(predict_coupon_usage(jack)[0]))
      print("Probability of Response from Jill = {:.4f}".
            ↪format(predict_coupon_usage(jill)[0]))

```

Probability of Response from Jack = 0.3944

Probability of Response from Jill = 0.3347

Jack is more likely to respond.

## 0.3 B-3:

After looking at the logistic regression model results, a few different indicators were found for rolling out coupon predictions to a large customer database...

# 1 The model coefficients:

BETA0 (or constant term): -2.0067 BETA1 (coeff. For X1 ): 0.3299 BETA2 (coeff. For X2): 0.9179

## 2 Odds Ratios:

X1: 1.3908 (Spending) X2: 2.5040 (Card Ownership)

---

These both show that spending and card ownership positively correlate with coupon usage, with card ownership having a stronger correlation. We should consider that the model predicted probabilities of a response of 0.3944 for Jack and 0.3347 for Jill when trying to figure out a cutoff probability. We think that a possible cutoff of 0.35 might be a good number to start off at because it would classify Jack as likely to use the coupon and Jill as unlikely.

But in order to actually make the cutoff as beneficial as it should be, there are a few steps that we need to start with. This includes splitting our data into training and test sets. Then, along with trial and error in picking cutoff points, we should use confusion matrices to see how accurate the predictions may be. In the end, the cutoff we choose should be found to balance the business goals we are trying to achieve and take into account coupon costs and benefits as well.

[ ]:

[ ]: