BANA 277 Homework #3

Group 10A

Chia-Jo Chen, Darshana Daga, Biying (Candice) Han, Casey Kan, Dong Wook Kim

1. Is online advertising effective for Star Digital? In other words, is there a difference in conversion rate between the treatment and control groups?

```
Call:
glm(formula = purchase ~ test, family = "binomial", data = sd_data)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.186  -1.186   1.169   1.169   1.202

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.05724    0.03882  -1.474   0.1404
test         0.07676    0.04104   1.871   0.0614 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
        Welch Two Sample t-test

data:  sd_data$purchase by sd_data$test
t = -1.8713, df = 3309.2, p-value = 0.06139
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.039289257  0.000916332
sample estimates:
mean in group 0 mean in group 1
      0.4856928       0.5048792
```

Overall, the estimate of the test is positive, so it seems that online advertising is effective for Star Digital. However, the p-value of the test is 0.0614, which is larger than 0.05, so it is not statistically significant at 95% confidence level. We cannot reject the null hypothesis of no difference between the averages of the two groups.

The intercept of purchase is -0.05724 and the test has a positive interaction with people purchasing the offering from Star Digital. If a customer is in the control group there is no change in purchase behaviour whereas if they are in the test group and are exposed to the advertisement by Star, the likelihood of purchase becomes positive 0.01952 (-0.05724+0.07676) or 1.9%.

2. Is there a frequency effect of advertising on purchase? In particular, the question is whether increasing the frequency of advertising (number of impressions) increases the probability of purchase?

```
Call:
glm(formula = purchase ~ Timp, family = binomial(), data = new_SD_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7410  -1.1265   0.1408   1.2166   1.2418

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.179392   0.014583  -12.30   <2e-16 ***
Timp         0.029201   0.001294   22.56   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35077  on 25302  degrees of freedom
Residual deviance: 34212  on 25301  degrees of freedom
AIC: 34216

Number of Fisher Scoring iterations: 5

> exp(coef(fit_Timp))
(Intercept)        Timp
  0.8357778   1.0296319
```

From the outcome above, it shows that the total impressions have a statistically significant result on purchase with $p < 0.05$. There is a positive effect of total impressions on purchase.

The exponential coefficient of total impressions is 1.0296319. In this case if a person views it as 1 impression , then the odds of the customer purchasing Star Digital’s offering increases by 1.03 times and makes the total purchasing’s odds ratio to 1.86 times each impression (intercept+coefficient).

```
Call:
glm(formula = purchase ~ imp_sum + test + imp_sum * test, family = binomial(),
    data = star)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.9145  -1.1266   0.1299   1.2156   1.2433

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.169577   0.042895  -3.953 7.71e-05 ***
imp_sum       0.015889   0.002876   5.524 3.32e-08 ***
test         -0.013903   0.045613  -0.305    0.761
imp_sum:test  0.015466   0.003207   4.823 1.42e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35077  on 25302  degrees of freedom
Residual deviance: 34190  on 25299  degrees of freedom
AIC: 34198

Number of Fisher Scoring iterations: 5

> exp(coef(fit_Timp_t))
 (Intercept)      imp_sum         test imp_sum:test
   0.8440214    1.0160156    0.9861932    1.0155865
```

Purchase = -0.1696+0.0159*Timp-0.0139*Test+0.0155*(Timp*Test)

1) For control group- People who don’t see ads from Star Digital, the test group is 0, the model in this case will be;

   $X_i = \beta_0 + \beta_1 * Impression\ count$

   Using the exponents in this formula gives us the odd ratio of purchase by control group;

   $X_i = -0.1696 + 0.0159 * Impression\ count$

   In this case with each impression count the odds of purchase increases by 0.0159 times and has the exponent exp(0.015889) = 1.0160156 or 1.6% odds of making purchase.

2) For the treatment group - People who see the advertisement by Star Digital, the test group is 1. For this case the model will be;

   $X_i = \beta_0 + \beta_1 * Impression\ Count + \beta_2 * test(=1) + \beta_3 * Impression\ Count * test(=1)$

$$X_i = -0.1696 + 0.0159 * impression\ count + (-0.0139) * 1 + 0.0155 * impression\ count * 1$$

In this case with each impression count the odds of purchase increases by (0.0159 + 0.0155) = 0.0314 . Taking exponent exp(0.0314) = 1.0314 or 3.1% odds of making a purchase.

Comparing the effect of control group (1.6%) and treatment group (3.1%) helps us come to a conclusion that showing ads to people has positive effects on making purchases. The more the frequency of ads the odds increases as well.

3. How does the conversion effectiveness of Sites 1-5 compare with that of Site 6?

```
glm(formula = purchase ~ test + imp_6 + imp_15 + imp_6 * test +
    imp_15 * test, family = binomial(), data = star)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.1280  -1.1195   0.1185   1.2217   1.2472

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.166556   0.042533  -3.916 9.01e-05 ***
test        -0.006087   0.045314  -0.134 0.893139
imp_6        0.003978   0.004294   0.927 0.354179
imp_15       0.019452   0.003443   5.650 1.61e-08 ***
test:imp_6   0.013483   0.005405   2.494 0.012616 *
test:imp_15  0.014617   0.003794   3.852 0.000117 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35077  on 25302  degrees of freedom
Residual deviance: 34166  on 25297  degrees of freedom
AIC: 34178

Number of Fisher Scoring iterations: 5

> exp(coef(model_3))
(Intercept)        test       imp_6      imp_15  test:imp_6 test:imp_15
  0.8465755   0.9939313   1.0039860   1.0196422   1.0135741   1.0147240
```

The result for the conversion effectiveness of sites 1-5 on purchase is statistically significant with a p-value < 0.05. There is a positive effect of sites 1-5 on purchase. The exponential coefficient of sites 1-5 is 1.0196422. This means that when 1 unit of sites 1-5 increases, the odds of purchasing Star Digital increases by 1.02 times.

The result for the conversion effectiveness of site 6 on purchase is statistically significant with a p-value < 0.05. There is a positive effect of sites 6 on purchase. The exponential coefficient of sites 1-5 is 1.0039860. This means that when 1 unit of sites 6increases, the odds of purchasing Star Digital increases by 1.004 times.

Purchase =
-0.1666+0.0040*Imp6+0.0195*Imp15-0.0061*Test+0.0135*(Imp6*Test)+0.0146*(Imp15*Test)

1) For control group- People who don't see add from Star Digital, the test group is 0, the model in this case will be;

   $Xi = \beta_0 + \beta_1 * Impression6 + \beta_3 * Impression15$

   Using the coefficient in this formula gives us the odd ratio of purchase by control group;

   $X_i =- 0.1666 + 0.003978 * Impression\ 6 + 0.0195 * \ Impression15$

   For site 1 to 5 odds within control groups of making a purchase increases by 0.0195 taking exponent exp(0.0195) odds of purchase are 1.0196 or 1.9%.

   For site 6 odds within the control group of making a purchase increases by 0.003978 , taking exponent exp(0.003978) odds of purchase are 1.0039860 or 0.39%.

   Based on the above numbers the control group performed better on websites 1 to 5.

2) For the treatment group - People who see the advertisement by Star Digital, the test group is 1. For this case the model will be;

   $$(X_i = \beta_0 + \beta_1 * Impression6 + \beta_2 * test(= 1) + \beta_3 * Impression15 + \beta_4 * Impression\ 6 * test(= 1) + \beta_5 * Impression15 * test(= 1))$$

   $$X_i =- 0.1666 + 0.003978 * Impression\ 6 + (- 0.006087) * test + 0.019452 * Impression15 + 0.013483 * Impression\ 6 * test + 0.014617 * Impression15 * test$$

   For site 1 to 5 odds within treatment groups of making a purchase increases by $(0.019452 + 0.014617) = 0.0340\ taking\ exp(0.0340) = 1.0340$ or 3.4%.

   For site 6 odds within the control group of making a purchase increases by (0.003978 + 0.013483) = 0.0175 taking exp(0.0175) = 1.0175 or 1.75%.

   Based on the above numbers the treatment group performed better on websites 1 to 5. In conclusion the website 1 to 5 shows better performance in terms of the conversion effectiveness.

4. (Optional) Which sites should Star Digital advertise on? In particular, should it put its advertising dollars in Site 6 or in Sites 1 through 5?

Offset = ln((1-PCR)/PCR) / ((1-SCR)/SCR)

$$= \ln\frac{(1-0.00153)/0.00153}{(1-0.5)/0.5}$$

```
Call:  glm(formula = purchase ~ imp_6, family = "binomial", data = new_SD_data,
    offset = offset)

Coefficients:
(Intercept)        imp_6
   -6.50210      0.01983

Degrees of Freedom: 25302 Total (i.e. Null);  25301 Residual
Null Deviance:      35080
Residual Deviance: 35010        AIC: 35020
```

```
Call:  glm(formula = purchase ~ imp_15, family = "binomial", data = new_SD_data,
    offset = offset)

Coefficients:
(Intercept)       imp_15
   -6.62572      0.03244

Degrees of Freedom: 25302 Total (i.e. Null);  25301 Residual
Null Deviance:      35080
Residual Deviance: 34220        AIC: 34220
```

```
Call:
glm(formula = purchase ~ imp_6 + imp_6 * test + imp_15 + imp_15 *
    test, family = binomial(), data = new_SD_data, offset = new_SD_data$offset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.1280  -1.1195   0.1185   1.2217   1.2472

Coefficients:
             Estimate Std. Error  z value Pr(>|z|)
(Intercept) -6.646556   0.042533 -156.267  < 2e-16 ***
imp_6        0.003978   0.004294    0.927 0.354179
test        -0.006087   0.045314   -0.134 0.893139
imp_15       0.019452   0.003443    5.650 1.61e-08 ***
imp_6:test   0.013483   0.005405    2.494 0.012616 *
test:imp_15  0.014617   0.003794    3.852 0.000117 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35077  on 25302  degrees of freedom
Residual deviance: 34166  on 25297  degrees of freedom
AIC: 34178

Number of Fisher Scoring iterations: 5

> exp(coef(model_4))
(Intercept)       imp_6        test      imp_15  imp_6:test test:imp_15
0.001298487 1.003986035 0.993931343 1.019642178 1.013574112 1.014724032
```

- The odds of purchase for customer in the treatment group visiting website 1-5;
  $\beta_3 * imp_{15} * impressions + \beta_5 * test * impression$

  = 0.019452 * 1 + 0.014617*1 = 0.034069 -> exp(0.034069) = 1.034656 odds

  Probability of purchase in this case (odds/(1+odds) = (1.034656/1+1.034656) = 0.5085164 or <u>50.85%</u>

- The odds of purchase for customer in the treatment group visiting website 6;

  $\beta_1 * imp_6 * impressions + \beta_4 * test * impressions$

  = 0.00398 * 1 + 0.01348*1 = 0.01746 -> exp(0.01746) = 1.017613 odds

  Probability of purchase in this case (odds/(1+odds) = (1.017613/1+1.017613) = 0.5043648 or <u>50.44%</u>

The customer lifetime value and cost of advertisements are as follows;

- Customer Lifetime Value = $1200
- Cost of advertising in website 1-5 per 1000 impression = $25
- Cost of advertising in website 6 per 1000 impression = $20
- Average impressions on website 1-5 = 6.091926
- Average impressions on website 6 = 1.783464

The ROI for advertising in website 1-5 is;

(CLV * Probability 1-5Cost of advertising * Average impressions) / Cost of advertising * Average impressions

- ➔ ($1200 * 0.5085164)- ($25/1000)*6.091926 /($25/1000)*6.091926
- ➔ 610.07 /0.1523 = $4005.71

The ROI for advertising in website 6 is;

(CLV * Probability 6 - Cost of advertising * Average impressions) / Cost of advertising * Average impressions

- ➔ ($1200 * 0.5043648) - ($20/1000)*1.783464 / ($20/1000)*1.783464
- ➔ 605.21 / 0.0357 = 16952.67$

The return on advertising on website 6 seems to give higher return on investment, a major reason could be that the average impression to convert into a purchase is much lower on website 6 compared to website 1 to 5. Hence, the company should invest in advertising on website 6.