# Hw2_Q4

Hongkai Lou

2024-10-04

```r
part <- read.csv('part.csv')
head(part)
```

```
##   tx wc          x  y
## 1  1  4   6.854164  3
## 2  1 32  29.893616  2
## 3  0  0 108.425476  0
## 4  1 27  63.583313  0
## 5  0  0  54.541656 84
## 6  0  0  23.914886  5
```

```r
dim(part)
```

```
## [1] 14178     4
```

```r
summary(part)
```

```
##        tx            wc               x                  y
##  Min.   :0.0   Min.   :  0.000   Min.   :  0.0233   Min.   :   0.0
##  1st Qu.:0.0   1st Qu.:  0.000   1st Qu.:  7.5445   1st Qu.:   0.0
##  Median :0.5   Median :  0.500   Median : 20.3936   Median :   6.0
##  Mean   :0.5   Mean   :  8.626   Mean   : 36.2646   Mean   :  55.1
##  3rd Qu.:1.0   3rd Qu.: 13.000   3rd Qu.: 46.7487   3rd Qu.:  45.0
##  Max.   :1.0   Max.   :189.000   Max.   :489.1250   Max.   :8188.0
```

(a) and (b)

```r
lm1 <- lm(log(y+1)~log(x+1)+tx, data = part )
lm2 <- lm(log(y+1)~log(x+1)+tx+log(wc+1), data = part)
summary(lm1)
```

```
##
## Call:
## lm(formula = log(y + 1) ~ log(x + 1) + tx, data = part)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6845 -1.2918 -0.0937  1.3063  6.1629
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.31705    0.04155  -7.631 2.47e-14 ***
## log(x + 1)   0.80318    0.01205  66.657  < 2e-16 ***
## tx           0.24438    0.02845   8.591  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.693 on 14175 degrees of freedom
## Multiple R-squared:  0.2406, Adjusted R-squared:  0.2405
## F-statistic:  2246 on 2 and 14175 DF,  p-value: < 2.2e-16
```

```r
summary(lm2)
```

```
##
## Call:
## lm(formula = log(y + 1) ~ log(x + 1) + tx + log(wc + 1), data = part)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6819 -1.2885 -0.0959  1.2999  6.1015
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.30823    0.04166  -7.398 1.46e-13 ***
## log(x + 1)   0.80026    0.01209  66.168  < 2e-16 ***
## tx           0.05039    0.07657   0.658  0.51053
## log(wc + 1)  0.07382    0.02706   2.729  0.00637 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.693 on 14174 degrees of freedom
## Multiple R-squared:  0.241,  Adjusted R-squared:  0.2409
## F-statistic:  1500 on 3 and 14174 DF,  p-value: < 2.2e-16
```

(c) Our Null Hypothesis is $B_2 = 0$, where participation does not have a significant effect on spending

Our Alternative Hypothesis is $B_2 \neq 0$, where participation has a significant effect on spending

Based on Model 1, the t-value of tx (participation) is 8.591, with the corresponding p-value less than 2e-16. our p-value < 2e-16. So that means participation has a significant effect on future spending. The estimate is 0.244, meaning that on average, if the customer participated, the amount spent by each customer in the week following the contest increase by 0.244.

(d) Approximately 1.28 greater.

```r
exp(0.24438)
```

```
## [1] 1.276829
```

(e)

```r
summary(lm2)
```

```
##
## Call:
## lm(formula = log(y + 1) ~ log(x + 1) + tx + log(wc + 1), data = part)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6819 -1.2885 -0.0959  1.2999  6.1015
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.30823    0.04166  -7.398 1.46e-13 ***
```

```
## log(x + 1)   0.80026    0.01209  66.168  < 2e-16 ***
## tx           0.05039    0.07657   0.658  0.51053
## log(wc + 1)  0.07382    0.02706   2.729  0.00637 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.693 on 14174 degrees of freedom
## Multiple R-squared:  0.241,  Adjusted R-squared:  0.2409
## F-statistic:  1500 on 3 and 14174 DF,  p-value: < 2.2e-16
```

We can imagine the value of tx will be highly impacted in model 2 because tx itself is an indicator that tx = (wc>0). When we include log(wc+1) in the model, it will have similar or more dramatic impact on the model than its indicator variable. log(wc+1) will be equal to 0 when wc = 0, and thus it can include more information and explain more variance, thus the effect of tx variable will be reduced.

(f) We see participation(tx) loses significance. Unlike model 1, participation has significant effect, once cognitive elaboration is included, effect of participation diminishes, with large p-value and will not pass the hypothesis test ($0.51 > 0.05$) at the 5% level. Whereas word count, with a coefficient of 0.07382, is more significant, affects the spending of customer in post-contest period, and will pass the hypothesis test at 5% level ($0.00637 < 0.05$).

(g): This suggests customers who put more cognitive effort into the entries tend to spend more in post-contest period. If the customer participate, and has high word count, the customer will have high cognitive elaboration, meaning that they are motivated and willing to spend more in the future.

(h) The results suggest that when designing future social media contests, the company should focus on encouraging deeper cognitive engagement rather than simple participation or writing just one word. Word count, a measure of engagement, was a strong predictor of future spending, while participation alone was not significant. Higher word count meaning more cognitive elaboration. Therefore, contests should require more thoughtful or detailed submissions, such as essays or creative content, to drive post-contest spending. Additionally, the company should target loyal customers, as their pre-contest spending is a good indicator of future behavior. Incentivizing higher levels of engagement through skill-based or interactive contests could further enhance future spending.