AB Testing Example

Diego Vallarino

2022-03-18

Contents

This is an exercise on how to perform univariate impact analysis (AB Testing).

To develop this analysis, certain variables were surveyed over a certain period of time, which are detailed below:

Pageviews: Number of unique cookies to view the course overview page that day. **Clicks**: Number of unique cookies to click the course overview page that day. **Enrollments**: Number of user-ids to enroll in the free trial that day. **Payments**: Number of user-ids who who enrolled on that day to remain enrolled for 14 days and thus make a payment.

In our analysis we will build a new attribute which will show the monetary value generated by each user.

We define the variable as: **Numbers of user-ids that enroll and then make a payment**

NOTE: *this is for money*

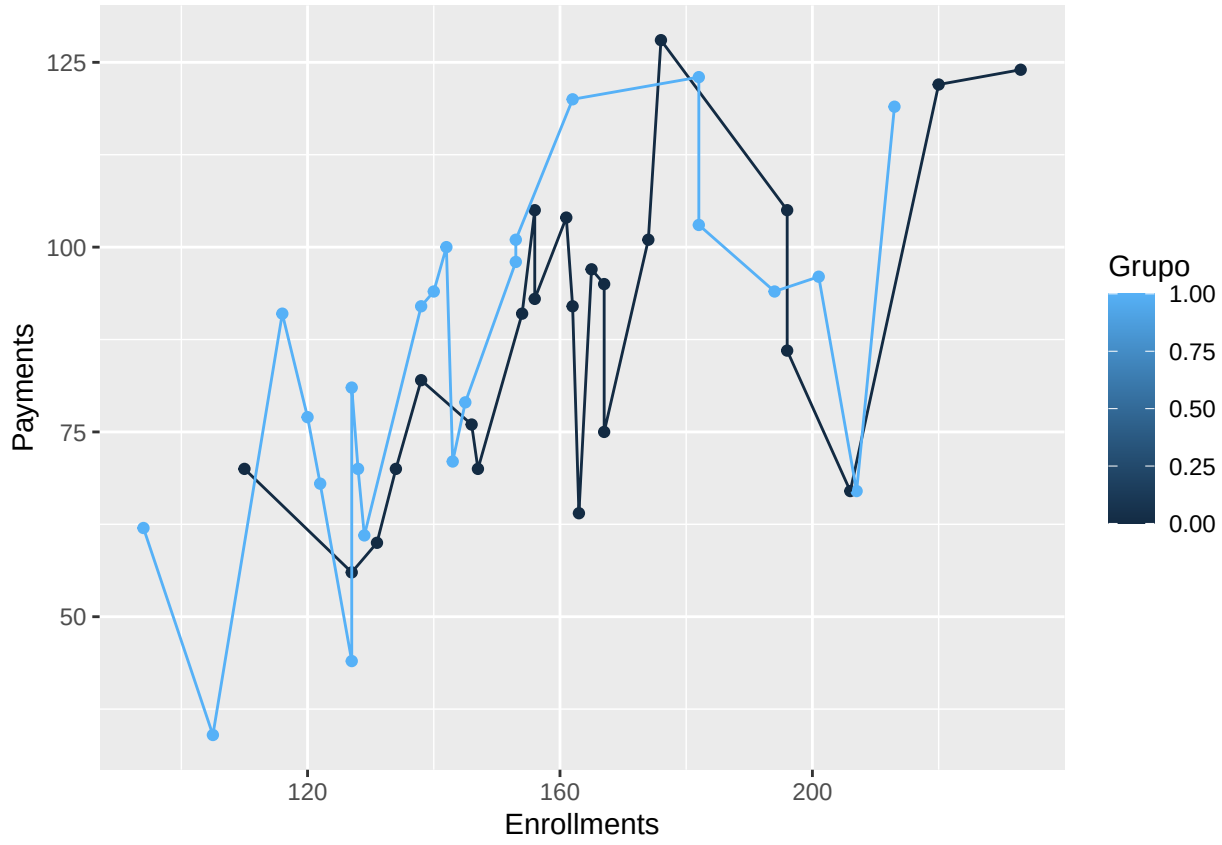
1 Preliminary analysis

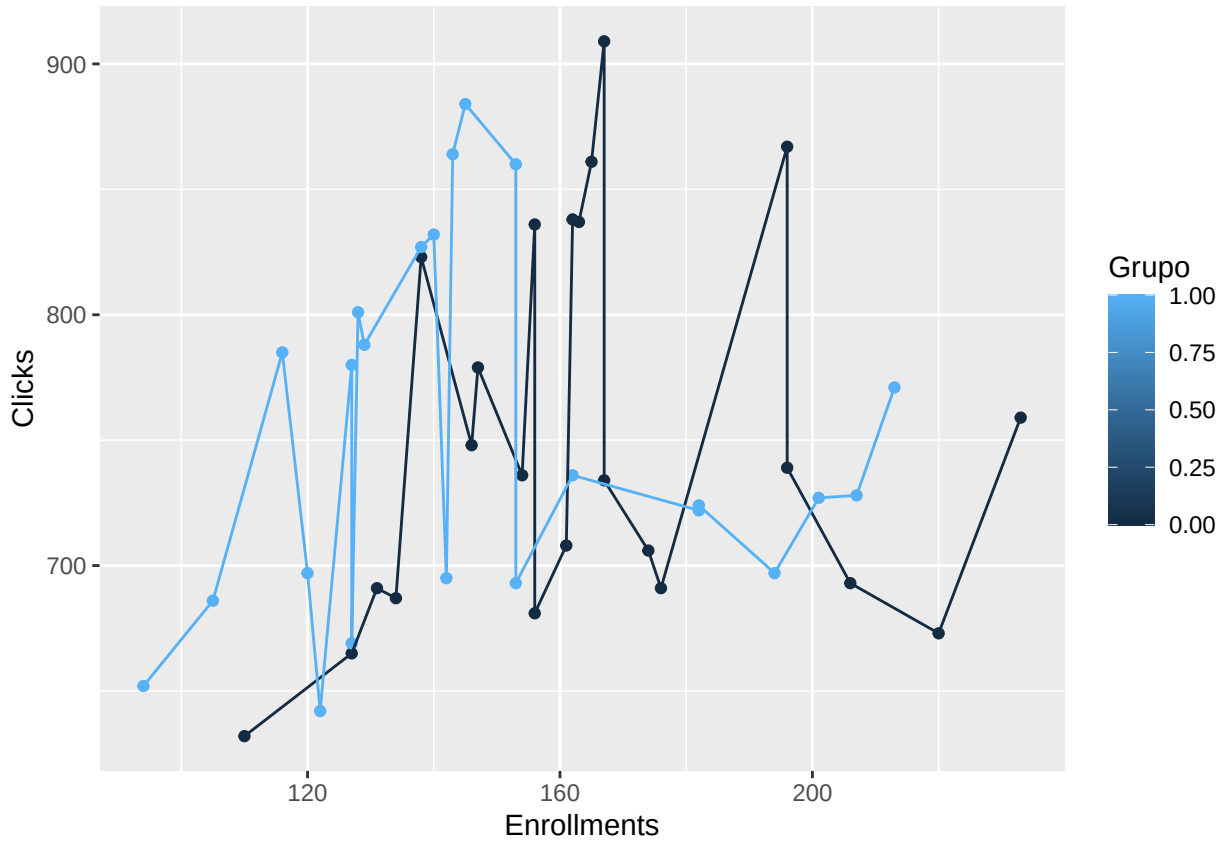
```
describe(Cont)
```

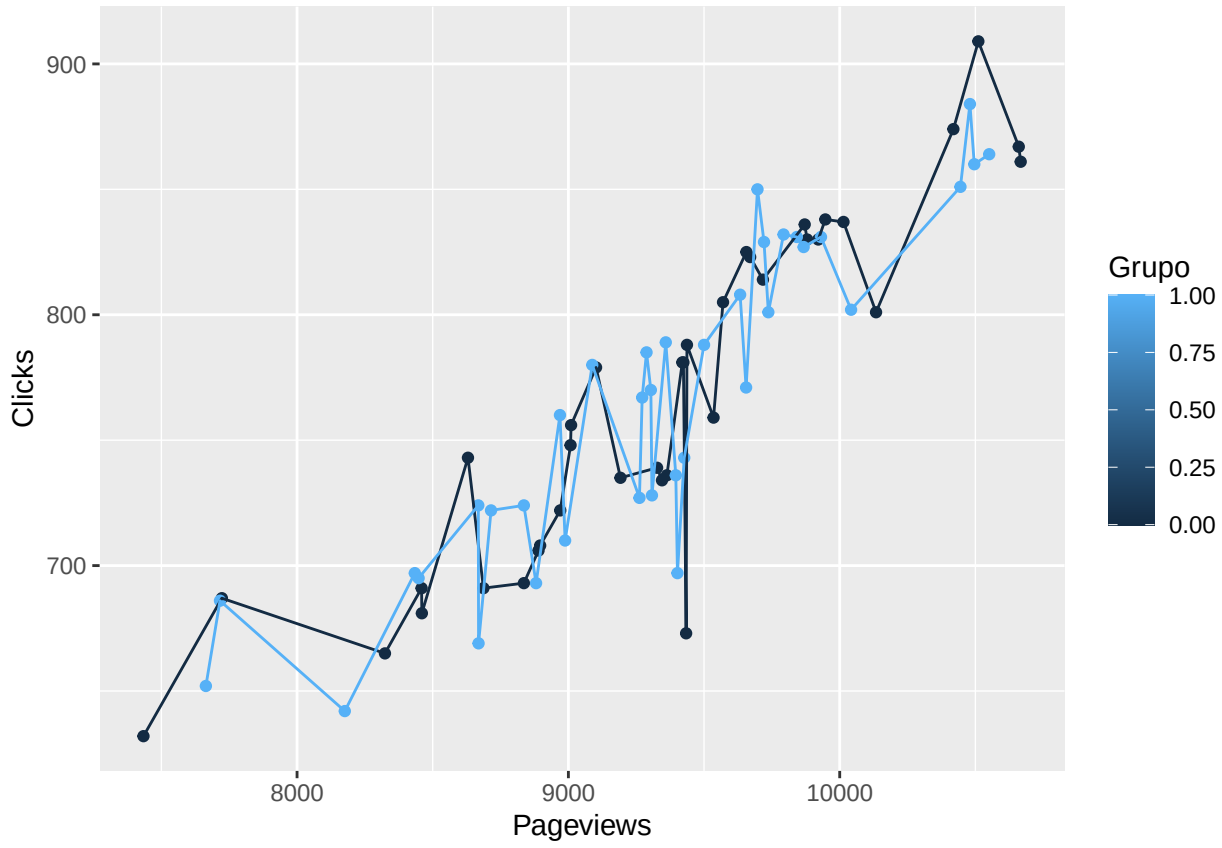
```
##             vars  n    mean     sd  median trimmed    mad     min      max
## Date*          1 37   19.00  10.82   19.00   19.00  13.34    1.00    37.00
## Pageviews      2 37 9339.00 740.24 9420.00 9362.06 682.00 7434.00 10667.00
## Clicks         3 37  766.97  68.29  759.00  766.39  94.89  632.00   909.00
## Enrollments    4 23  164.57  29.98  162.00  162.89  22.24  110.00   233.00
## Payments       5 23   88.39  20.65   91.00   87.63  22.24   56.00   128.00
## ratepay        6 23    0.54   0.09    0.55    0.54   0.06    0.33     0.73
##               range  skew kurtosis     se
## Date*          36.0  0.00    -1.30   1.78
## Pageviews    3233.0 -0.36     0.00 121.69
## Clicks        277.0  0.05    -1.02  11.23
## Enrollments   123.0  0.50    -0.31   6.25
## Payments       72.0  0.28    -1.01   4.31
## ratepay         0.4 -0.25    -0.40   0.02
```

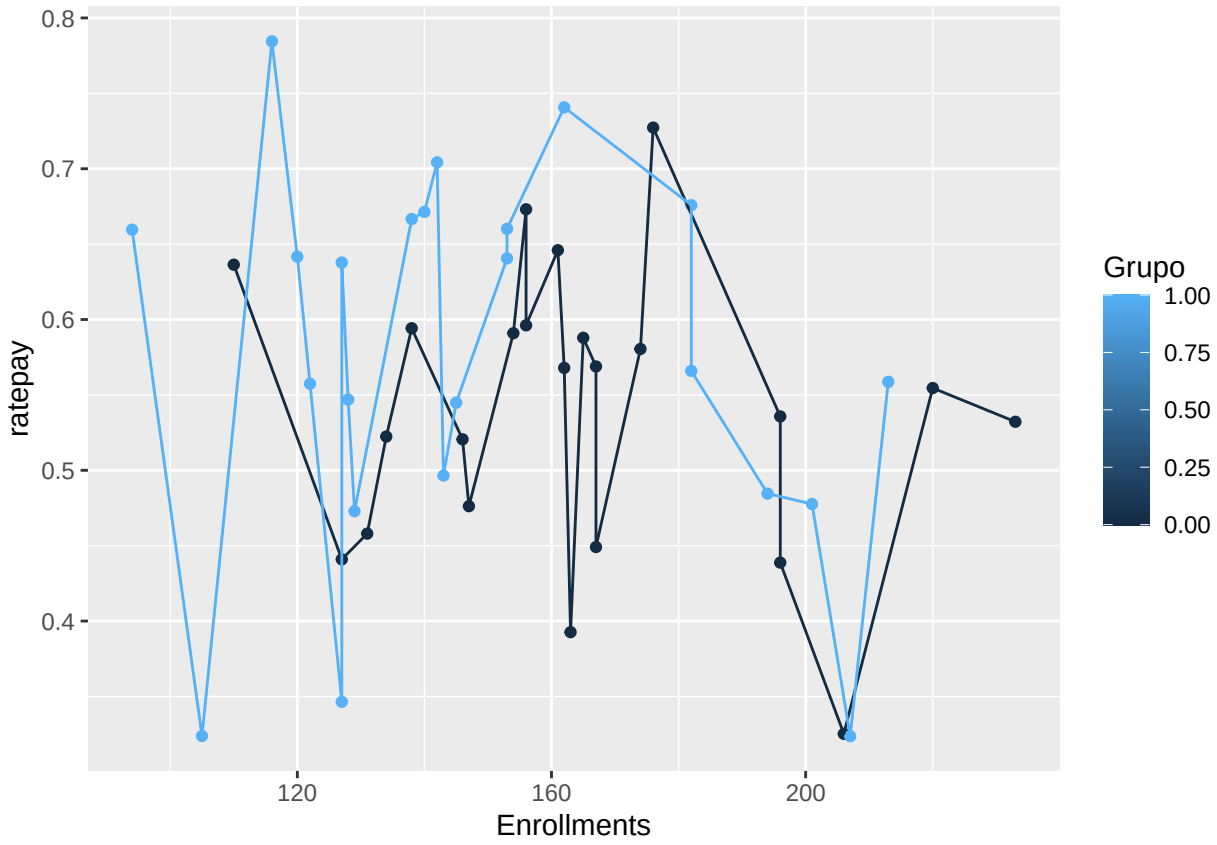
```
describe(Exp)
```

```
##              vars  n    mean     sd  median trimmed    mad     min      max
## Date*           1 37   19.00  10.82   19.00   19.00  13.34    1.00    37.00
## Pageviews       2 37 9315.14 708.07 9359.00 9341.19 643.45 7664.00 10551.00
## Clicks          3 37  765.54  64.58  770.00  766.26  84.51  642.00   884.00
## Enrollments     4 23  148.83  33.23  142.00  147.58  29.65   94.00   213.00
## Payments        5 23   84.57  23.06   91.00   85.47  20.76   34.00   123.00
## ratepay         6 23    0.57   0.13    0.57    0.58   0.14    0.32     0.78
##                range  skew kurtosis     se
## Date*          36.00  0.00    -1.30   1.78
## Pageviews    2887.00 -0.35    -0.21 116.41
## Clicks        242.00 -0.04    -1.11  10.62
## Enrollments   119.00  0.48    -0.92   6.93
## Payments       89.00 -0.26    -0.61   4.81
## ratepay         0.46 -0.51    -0.67   0.03
```




900
800
700
Clicks
120
160
200
Enrollments
Grupo
1.00
0.75
0.50
0.25
0.00


900
800
700
Clicks
8000
9000
10000
Pageviews
Grupo
1.00
0.75
0.50
0.25
0.00



2 Hypothesis test on equality of means

```
##
##  Welch Two Sample t-test
##
## data:  Exp$ratepay and Cont$ratepay
## t = 1.0081, df = 40.505, p-value = 0.3194
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03347511  0.10016013
## sample estimates:
## mean of x mean of y
## 0.5731398 0.5397973
```

```
##
##      Two-sample t test power calculation
##
##               n = 37
##           delta = 0.03
##              sd = 1
##       sig.level = 0.05
##           power = 0.03342822
##     alternative = two.sided
##
```

```
## NOTE: n is number in *each* group

##
##      Two-sample t test power calculation
##
##               n = 17442.92
##           delta = 0.03
##              sd = 1
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

The experiment is not applicable because we do not see effects in the ratio that we consider, and also it does not have the sample power to be able to be inferential to the population.

NOTE: A logistic regression analysis will be performed with the ultimate goal of demonstrating that this method was not feasible in this exercise. The model does not meet the conditions to use this method

```
##
## Call:
## lm(formula = ratepay ~ ., data = df[, -1])
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.073547 -0.010378 -0.002479  0.007705  0.064609
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.413e-01  4.304e-02  12.576 1.76e-15 ***
## Pageviews    1.043e-05  1.486e-05   0.702    0.487
## Clicks      -1.302e-04  1.534e-04  -0.849    0.401
## Enrollments -3.352e-03  1.860e-04 -18.022  < 2e-16 ***
## Payments     6.244e-03  2.225e-04  28.063  < 2e-16 ***
## Grupo        4.650e-03  7.828e-03   0.594    0.556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 0.02518 on 40 degrees of freedom
##   (28 observations deleted due to missingness)
## Multiple R-squared: 0.9552,
## Adjusted R-squared: 0.9496
## F-statistic: 170.6 on 5 and 40 DF,  p-value: < 2.2e-16
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lm$residuals
## W = 0.90729, p-value = 0.001413
```

```
##
##  studentized Breusch-Pagan test
```

```
##
## data:  lm
## BP = 15.289, df = 5, p-value = 0.009196
```