

A/B test results

Jonathan P. Latner, PhD

October 11, 2022

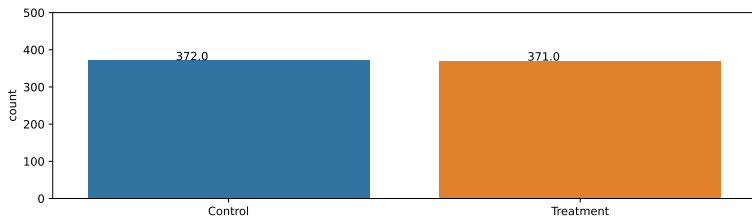
Overview

- A/B test conducted during the month January, 2019
- Sample size 743 (\cong 24 individuals per day)
- Results suggest no effect of treatment on outcomes

A/B test sample

Size = 743 (even split by treatment)

Figure 1: Treatment

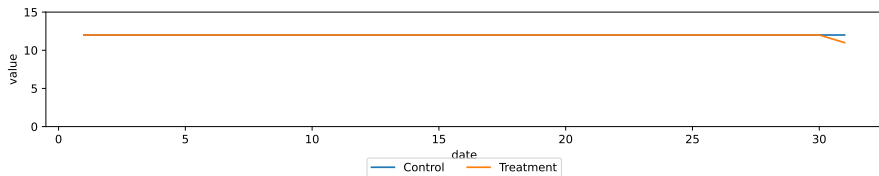


Time period

One month, January 2019

\approx 12 individuals per day

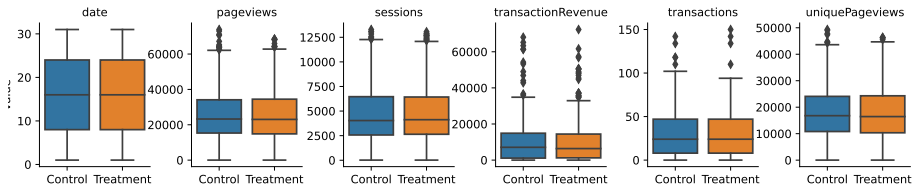
Figure 2: Treatment by date



Descriptives - numerical variable

No difference between treatment and control

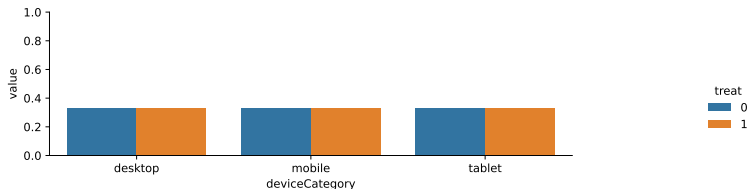
Figure 3: Numerical variables by treatment



Descriptives - categorical variable

No difference between treatment and control

Figure 4: Categorical variable (device) by treatment



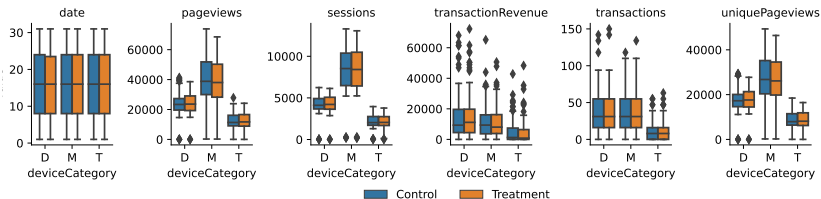
Descriptives - crosstab numerical and control

Difference between devices

- Mobile has higher pageviews, uniqeviews, and sessions
- Tablet has lower revenues and transactions
- Desktop has slightly higher revenues than mobiles

But, no difference between treatment and control

Figure 5: Numerical variables by device and treatment

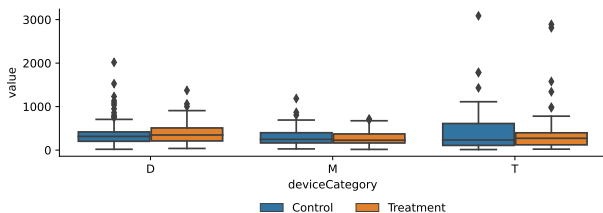


What about a new variable?

Avg. revenue per transaction = revenue/transaction

Among tablet users, treatment appears to reduce outliers, but not average

Figure 6: Average revenue per transaction by device and treatment

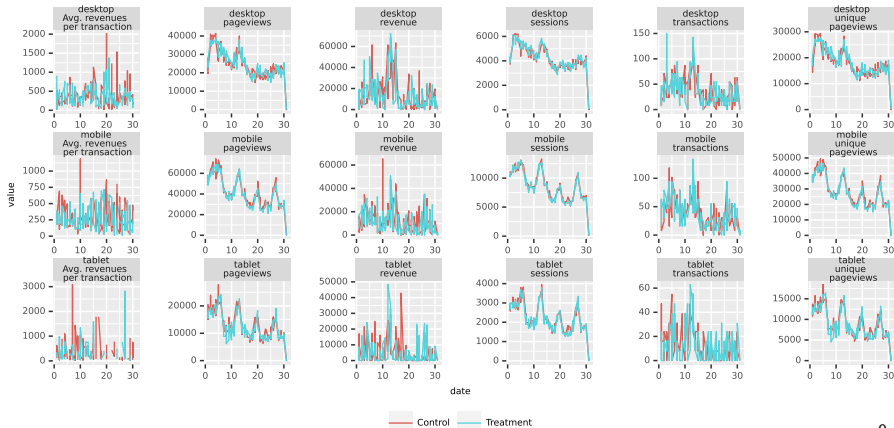


Time trends

Variables are declining over time by device category

No difference trends over time by treatment

Figure 7: Numerical variables over time by device and treatment



Regressions

Model 0 only has main effect of device category and treatment

Model 1 adds an interaction between device category and treatment

Regression models confirm descriptive statistics

There is a main effect of Device category, but no effect of treatment

Table 1: Parameter estimates of effect of device and treatment on outcomes

	Sessions		Pageviews		Unique views		Transactions		Revenue		Avg. Rev per Trns	
	0	1	0	1	0	1	0	1	0	1	0	1
deviceCategory[T.mobile]	4199.86** (157.37)	4292.09** (222.58)	15819.55** (919.92)	16562.90** (1300.84)	9521.38** (625.99)	10014.29** (885.21)	1.53 (2.09)	1.53 (2.95)	-3324.78** (1012.79)	-2573.68+ (1432.07)	-89.01** (29.28)	-69.71+ (41.51)
deviceCategory[T.tablet]	-2032.85** (157.37)	-1984.10** (222.58)	-11734.81** (919.92)	-11357.18** (1300.84)	-9163.58** (625.99)	-8876.89** (885.21)	-25.37** (2.09)	-24.65** (2.95)	-9641.13** (1012.79)	-8884.33** (1432.07)	32.53 (32.29)	52.02 (45.01)
treat[T.Treatment]	-14.66 (128.45)	79.70 (223.03)	-275.55 (750.86)	474.80 (1303.48)	-101.01 (510.95)	420.84 (887.01)	0.28 (1.70)	0.76 (2.96)	-36.62 (826.66)	972.73 (1434.98)	-15.13 (25.20)	9.49 (41.60)
deviceCategory[T.mobile]:treat[T.Treatment]		-184.83 (315.09)		-1489.73 (1841.53)		-987.93 (1253.15)		0.00 (4.18)		-1506.28 (2027.31)		-38.45 (58.64)
deviceCategory[T.tablet]:treat[T.Treatment]		-97.86 (315.09)		-758.30 (1841.53)		-575.51 (1253.15)		-1.46 (4.18)		-1517.69 (2027.31)		-39.68 (64.70)
Intercept	4239.36** (128.45)	4192.37** (157.39)	24507.43** (750.86)	24133.77** (919.83)	18047.76** (510.95)	17787.89** (625.94)	37.21** (1.70)	36.97** (2.09)	14528.87** (826.66)	14026.24** (1012.63)	381.25** (24.35)	368.83** (29.54)
Obs	743	743	743	743	743	743	743	743	743	743	631	631
Adj. R ²	0.69	0.69	0.55	0.55	0.55	0.54	0.22	0.22	0.11	0.11	0.02	0.02

Regressions

Model 2 adds main effect of date

Model 3 adds an interaction with date

Regression models confirm descriptive statistics (declining trends)

Tablets trends are still negative: interaction with date < main effect

Table 2: Parameter estimates of effect of device and treatment on outcomes

	Sessions		Pageviews		Unique views		Transactions		Revenue		Avg. Rev per Trns	
	2	3	2	3	2	3	2	3	2	3	2	3
deviceCategory[T.mobile]	4207.06** (125.31)	6426.77** (254.32)	15865.33** (692.41)	26364.37** (1436.44)	9552.70** (469.26)	16280.17** (977.97)	1.60 (1.93)	4.92 (4.35)	-3304.55** (977.39)	-4210.15* (2225.91)	-89.01** (29.31)	-54.36 (66.46)
deviceCategory[T.tablet]	-2025.65** (125.31)	-2504.97** (254.32)	-11689.03** (692.41)	-15328.50** (1436.44)	-9132.27** (469.26)	-11817.11** (977.97)	-25.31** (1.93)	-36.17** (4.35)	-9620.90** (977.39)	-13531.43** (2225.91)	32.60 (32.35)	46.66 (70.90)
treat[T.Treatment]	-19.46 (102.28)	66.06 (226.34)	-306.07 (565.16)	463.26 (1278.38)	-121.88 (383.02)	392.07 (870.36)	0.24 (1.57)	-0.36 (3.87)	-50.11 (797.76)	951.94 (1980.99)	-15.14 (25.22)	0.46 (61.07)
date	-118.36** (5.72)	-84.71** (10.18)	-752.85** (31.63)	-628.79** (57.51)	-514.99** (21.44)	-444.58** (39.16)	-1.00** (0.09)	-1.20** (0.17)	-332.67** (44.65)	-462.84** (89.12)	0.08 (1.45)	0.06 (2.82)
deviceCategory[T.mobile]:treat[T.Treatment]		-174.53 (222.89)		-1412.55 (1258.89)		-933.52 (857.09)		0.14 (3.81)		-1449.56 (1950.79)		-38.45 (58.82)
deviceCategory[T.tablet]:treat[T.Treatment]		-87.56 (222.89)		-681.12 (1258.89)		-521.09 (857.09)		-1.32 (3.81)		-1460.97 (1950.79)		-39.22 (64.97)
deviceCategory[T.mobile]:date		-133.42** (12.48)		-612.59** (70.51)		-391.62** (48.00)		-0.21 (0.21)		102.28 (109.26)		-1.00 (3.38)
deviceCategory[T.tablet]:date		32.55** (12.48)		248.21** (70.51)		183.76** (48.00)		0.72** (0.21)		290.44** (109.26)		0.38 (3.73)
treat[T.Treatment]:date		0.21 (10.18)		-4.10 (57.51)		-1.60 (39.16)		0.06 (0.17)		-2.25 (89.12)		0.58 (2.90)
Intercept	6128.38** (137.14)	5547.69** (197.33)	36522.44** (757.80)	34194.39** (1114.52)	26266.68** (513.58)	24901.16** (758.79)	53.23** (2.11)	56.24** (3.37)	19838.16** (1069.69)	21431.66** (1727.06)	380.09** (32.92)	367.96** (52.21)
Obs	743	743	743	743	743	743	743	743	743	743	631	631
Adj. R ²	0.80	0.84	0.74	0.79	0.74	0.79	0.33	0.35	0.17	0.17	0.02	0.01

Questions

- What conclusions do you have?
 - ▶ The experiment suggests little effect of treatment on outcome variables
 - ▶ Whatever treatment effect exists is smaller than device category
- What does this mean for us and our test?
 - ▶ The treatment has no effect, or . . .
 - ▶ The treatment is too small to have an effect
- What should we do and how should we proceed?
 - ▶ Examine alternative outcome variables
 - ▶ Examine differences by demographic groups
- Is there anything specific with regards to device categories?
 - ▶ Mobile has highest pageviews, uniqueviews, and sessions
 - ▶ Tablet has lowest revenues and transactions
 - ▶ Desktop has slightly higher revenues than mobiles (but not sig.)
- Any other things you noticed or found interesting?
 - ▶ Time trends - declining
 - ▶ Mobile \approx Desktop in revenues and transactions

Thank you