Q4:

**Over the course of a week, I divided invites from about 300 requests among four new variations of the quote form as well as the baseline form we've been using for the last year. Here are my results:**

* **Baseline: 32 quotes out of 595 viewers**
* **Variation 1: 30 quotes out of 599 viewers**
* **Variation 2: 18 quotes out of 622 viewers**
* **Variation 3: 51 quotes out of 606 viewers**
* **Variation 4: 38 quotes out of 578 viewers**

**What's your interpretation of these results? What conclusions would you draw? What questions would you ask me about my goals and methodology? Do you have any thoughts on the experimental design? Please provide statistical justification for your conclusions and explain the choices you made in your analysis.**

**For the sake of your analysis, you can make whatever assumptions are necessary to make the experiment valid, so long as you state them.**

===============================================================================

I would ask you the following questions about your goals and methodology:

**1. Testing against baseline:**

**Would you like to test if each variation is significantly different than the baseline? Or would you like to test if each variation is ‘greater than’ or ‘less than’ the baseline?**

We define the variable Y = number of quotes out of total viewers who saw that version (Same for each other variation). Then Y follows a Bin (n, π0 )

32 quotes out of 595 viewers = 0.05378151

30 quotes out of 599 viewers

18 quotes out of 622 viewers

51 quotes out of 606 viewers

30 quotes out of 578 viewers

a) We represent the ‘not equal to’ test mathematically as having null hypothesis

H0 : π = π0 and alternative as HA : π ≠ π0. The score test tests the SE under the null hypothesis H0 . Where Z is the standard normal random variable.

Z =

where

The associated two-tailed P-value is given by calculating

1. the probability of the observed valued result under the null hypothesis

P = Pr( Y = y | H0 : π = π0 ) =

1. the proabilities of all other outcomes.

Pj =Pr( Y = j | H0 : π = π0 ) =

1. sum the probabilities Pj in (2) that are less than or equal to the observed probability P in (1):

P-value =

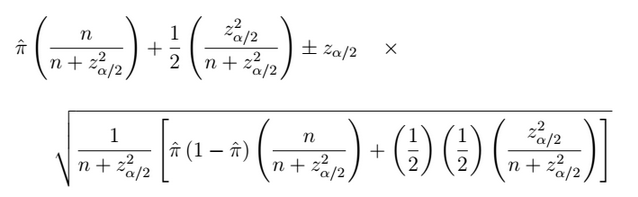
Note: special case when H0 : π0 = 0.5

p-value =

The P value is the sum of the probabilities for events that are at least or more extreme (in the direction of the alternative hypothesis) than what we observed, given that the null hypothesis H0 : π = π0 is true.

As a rule of thumb, p ≤ 0.05 means that we have very strong presemption against the null hypothesis and we can reject the null hypothesis since the observed result would be highly unlikely under the null hypothesis.

The Score Confidence Interval is computed by inverting the score test statistics. The resulting 100(1-α) score interval is



The R code for this 1-sample proportions test without continuity correctionis:

> prop.test(30, 599, p=0.05378151,correct=FALSE)

1-sample proportions test without continuity correction

data: 30 out of 599, null probability 0.05378151

X-squared = 0.161, df = 1, p-value = 0.6883

alternative hypothesis: true p is not equal to 0.05378151

95 percent confidence interval:

0.03530448 0.07059643

sample estimates:

p

0.05008347

We reject the null hypothesis when the p-value is less than α. In this case, the p-value is well over α = 0.05 and we fail to conclude that variation 1 brings a different result than the baseline.

b) We represent the ‘less than’ test mathematically as having null hypothesis

H0 : π = π0 and alternative as HA : π < π0

Then under H0 : π = π0, you would expect y ≈ n π0 and

Under HA : π < π0, you would expect y < n π0.  In other words, you would expect y to be ‘small’ under the alternative.

If you observe Y = y successes, the p-value is

Pr( Y≤ y | H0 : π = π0 ) =

Here’s the R code for testing whether variation 2 is signicifantly ‘worse’ than the baseline.

> prop.test(18, 622, p=0.05378151,alternative ='less',correct=FALSE)

1-sample proportions test without continuity correction

data: 18 out of 622, null probability 0.05378151

X-squared = 7.5433, df = 1, p-value = 0.003012

alternative hypothesis: true p is less than 0.05378151

95 percent confidence interval:

0.00000000 0.04219807

sample estimates:

p

0.02893891

Based on the p-value and the confidence interval, we can reject the null hypothesis and conclude that variation 2 brings in significantly less quotes than the baseline.

c) We represent the ‘greater than’ test mathematically as having null hypothesis

H0 : π = π0 and alternative as HA : π > π0

p-value : Pr( Y ≥ y | H0 : π = π0 ) =

An example of testing whether variation 3 brings in a proportionally higher number of quotes than the baseline:

> prop.test(51, 606, p=0.05378151,alternative ='greater',correct=FALSE)

1-sample proportions test without continuity correction

data: 51 out of 606, null probability 0.05378151

X-squared = 10.9884, df = 1, p-value = 0.0004584

alternative hypothesis: true p is greater than 0.05378151

95 percent confidence interval:

0.06740569 1.00000000

sample estimates:

p

0.08415842

Based on the p-value and the confidence interval, we can conclude that variation 3 brings in a proportionally higher number of quotes than the baseline.

**2. Testing among variations:**

**Would you like to test whether or not the results are statistically ‘different’,’greater than’, or ‘less than’ among variations?**

We can use 2-sample test for equality of proportions without continuity correction. For example, comparing variations 1 and 2, we define **y1** = **30 quotes under variation 1.**

**y2 = 18 quotes under variation 2. n0 = 599 viewers that saw variation 1. n1 = 622 viewers that saw variation 2. Then,**

The confidence interval is computed as:

The R code for such a test:

y <- c(30,18)

n <- c(599,622)

prop.test(y,n)

prop.test(y,n,alternative="two.sided",correct=F)

Output:

> prop.test(y,n,alternative="two.sided",correct=F)

2-sample test for equality of proportions without continuity correction

data: y out of n

X-squared = 3.6124, df = 1, p-value = 0.05735

alternative hypothesis: two.sided

95 percent confidence interval:

-0.0007337083 0.0430228397

sample estimates:

prop 1 prop 2

0.05008347 0.02893891

Based on the p-value and confidence interval we are able to reject the null hypothesis at α = 0.10. The proportions of quotes obtained when comparing the two variations (1 and 2) are statistically significantly different at α = 0.10.