q3

May 17, 2017

0.1 Question 3:

As context, here's a 10,000-foot view of the Acme Corp product:

0.2

- A consumer posts a request for a service needed. Every request is in some category (e.g., Catering, Personal Training, Interior Design) and some location (e.g., New York, San Francisco).
- We match the request up with appropriate service providers and send each of those providers an invite to quote on the request.
- Providers view the invite and some choose to send a quote to the consumer expressing interest.

For the following questions, please be as specific and thorough as possible in your answers, quantify your statements as much as you can, and explain your computations. Include code you used where appropriate. You're free to use any software you like; it's OK if we can't run the analysis ourselves. You're encouraged to be as technical as you like in your answers, they don't need to be accessible to general readers (though that's an important part of the actual job).

0.2.1 Split Test Analysis

0.3

I've just concluded a test of our quote form. After receiving an invite, service providers come to the quote form to view the consumer request and choose whether or not to pay to send a quote. My goal was to determine if certain changes to the design of the form would cause more providers to send a quote after coming to the page.

Over the course of a week, I divided invites from about 3000 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. So, for example, your response might follow the form "I would ask you A, B and C about your goals and methodology. Assuming the answers are X, Y and Z, then here's my analysis of the results... If I were to run it again, I would consider changing...".

Here is the data formatted as a CSV:

Bucket,Quotes,Views
Baseline,32,595
Variation 1,30,599
Variation 2,18,622
Variation 3,51,606
Variation 4,38,578

Q: What's your interpretation of these results? A: My interpretation of the A/B testing is that there area varying conversion rates for quotes per variation (5.38%, 5.00%, 2.89%, 8.42%, 6.62%) respectively.

Q: What conclusions would you draw? A: I would conclude nothing until answering the below questions.

Q: What questions would you ask me about my goals and methodology? A: I would ask the following: - Are these variations statistically significant? - Are these all unique viewers? - Were the users selected randomly who received a variation of the control? - Are the control population and variant population comprable? - Why Multi-invariants were not used? - Were the quotes higher or lower on average per variation? - Was there a difference in Clicks-to-Quotes? Meaning, did some variations make it easier to fill out a quote? - Is one week adequate time to adjust to changes?

Q: Do you have any thoughts on the experimental design? A: - Single-Metric design, could use more variables based on my above questions on methodology. - Was there any "Change Aversion" noticed or vocalized by the users?

Assumptions

1. Variations were presented randomly.
2. Quotes averages were the same.
3. Variation in form complexity and ease of use were neglegible.
4. Viewers were unique.
5. The quotes are independant, one quote or many quotes does not affect other viewers quoting.

Data

```
In [47]: import pandas

         data = pandas.DataFrame([{'Bucket': 'Baseline',
                                   'Quotes': 32,
                                   'Views': 595},
                                  {'Bucket': 'Variation 1',
                                   'Quotes': 30,
                                   'Views': 599},
                                  {'Bucket': 'Variation 2',
                                   'Quotes': 18,
                                   'Views': 622},
                                  {'Bucket': 'Variation 3',
                                   'Quotes': 51,
                                   'Views': 606},
                                  {'Bucket': 'Variation 4',
                                   'Quotes': 38,
                                   'Views': 578},
                                 ])

         print(data.describe())

           Quotes         Views
count    5.000000      5.000000
mean    33.800000    600.000000
std     12.049896     16.046807
min     18.000000    578.000000
25%     30.000000    595.000000
50%     32.000000    599.000000
75%     38.000000    606.000000
max     51.000000    622.000000
```

Hypothesis Testing

With such a relatively small difference in conversion rates (5.38%, 5.00%, 2.89%, 8.42%, 6.62%), however, can we convincingly say that the variation converts better?

To test the *statistical significance* of a result like this, a hypothesis testing can be used.

```
In [48]: # Take a look at the dataframe to remind of of the dataset
         population_quotes_mean = data['Quotes'].mean()
         population_views_mean = data['Views'].mean()

         print('Population Quotes Mean = {}'.format(population_quotes))
         print('Population Views Mean = {}\n'.format(population_views))

         print('Original DataFrame:')
         print(data)
```

3

```python
# Pull out the test by index location (.iloc[])
control = data.iloc[0]
variation_1 = data.iloc[1]
variation_2 = data.iloc[2]
variation_3 = data.iloc[3]
variation_4 = data.iloc[4]
```

```
Population Quotes Mean = 33.8
Population Views Mean = 600.0

Original DataFrame:
        Bucket  Quotes  Views
0      Baseline      32    595
1   Variation 1      30    599
2   Variation 2      18    622
3   Variation 3      51    606
4   Variation 4      38    578
```

First we would want to calculate the population mean and then the confidence intervals (95% in most cases). Then of each variation of the new forms, determine whether that variation of the forms p-value is greater or less than 0.05 giving us reason to accept of reject the 'Null Hypothesis' (if the sample data (variation(s)) are drawn for the same population).

After determining that we can infer whether or not each variation had a statistically significant increase, or decrease in quotes compared to the baseline.