The Split Test Analysis of Acme Corp Product's Quote Form

The below data represents the test results observed from dividing 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.

- **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

I would like to ask the below questions about the goals and methodology of the above conducted experiment:

1. What kind of split test was conducted? Multivariate or A/B test?
2. What was changed in Variation 1, 2, 3,4?
3. Did you run the control and variations at the same time?
4. When do you say the experiment was successful?
5. What is the accepted confidence level of the alternate hypothesis?

Assuming the answers are:

1. A/ B test.
2. The changes are:
    a. Variation 1: Change header color to blue
    b. Variation 2: Change "Pay to Quote" button color to none
    c. Variation 3: Change "Pay to Quote" button text to "Send Quote"
    d. Variation 4: Align "Pay to Quote" button to bottom of the page.
3. Yes. All the baseline and variations were run at the same time for same sample size. For instance, at a given time, 100 requests were tested with baseline design, 100 with Variation 1 ,100 with Variation 2 etc.…
4. The experiment is successful if we can get a conversion rate of 25% at this time.
5. The confidence level of the alternate hypothesis should be 95%.

Then, below is the analysis of the results.

For the sake of the analysis, I assume that quote form looks like below:

Form: baseline

Opportunity to Cater for a wedding in LA

Pay to Quote

Cancel

Customer Name	**George Anderson**
Address	**Irvine, California**
Customer Request	**I need a wedding caterer for my daughter's wedding. We prefer Italian food. The wedding is on Jan 1st, 2018 in Los Angeles. Our budget is 3000 USD. Send details of your quote and We will back to you in a day.**
Customer Id	**XK123KU**

I would like to assume that the button "Pay to Quote" in the above form leads to another page where in the provider is charged and the quote is submitted to the customer. The button "Cancel" will exit the form.

The above form will be assumed as the baseline. Now there are 4 variations to the form.

To test the effectiveness of the variation, we can use conversion rate.

$$\text{Conversion Rate \%} = \frac{\text{Quotes}}{\text{Views}} \times 100$$

$$\text{Difference in Conversion Rate} = (CR_B - CR_A) / CR_A$$

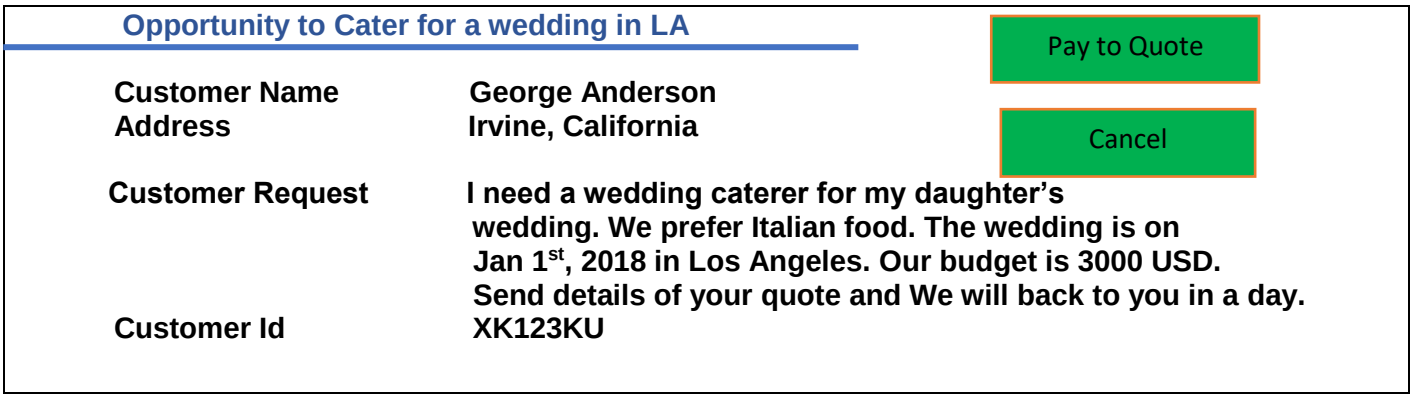
A Baseline Version
B Variation
$\mathbf{CR_B}$ Conversion Rate of Variation.
$\mathbf{CR_A}$ Conversion Rate of Control or baseline.

From the data provided in the problem, I know that the number of views is 595 and number of quotes sent is 32 for the baseline design (null hypothesis). From this one can find the conversion rate (the percentage of your visitors who end up reaching a given goal).

Conversion Rate = 5.38%.

Variation 1:

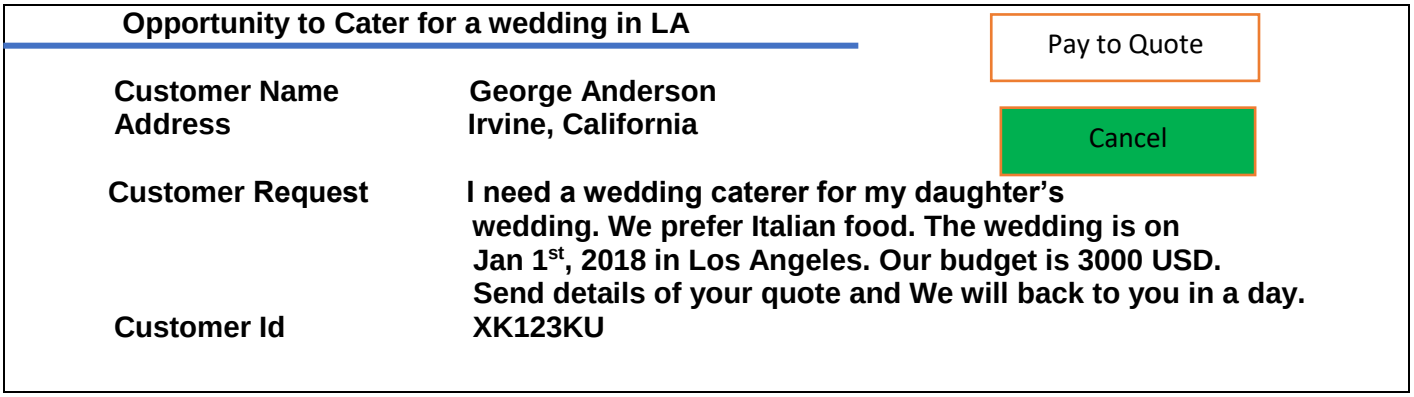
The assumed variation to the form is changing the color of the header text to blue.



Conversion rate with Variation 1 is 5.01 %. The test result is not significant. The observed difference in conversion rate (-6.88%) between baseline and this variation isn't big enough to declare a significant winner. There is no real difference in performance between A(Baseline) and B (Variation 1).

Variation 2:

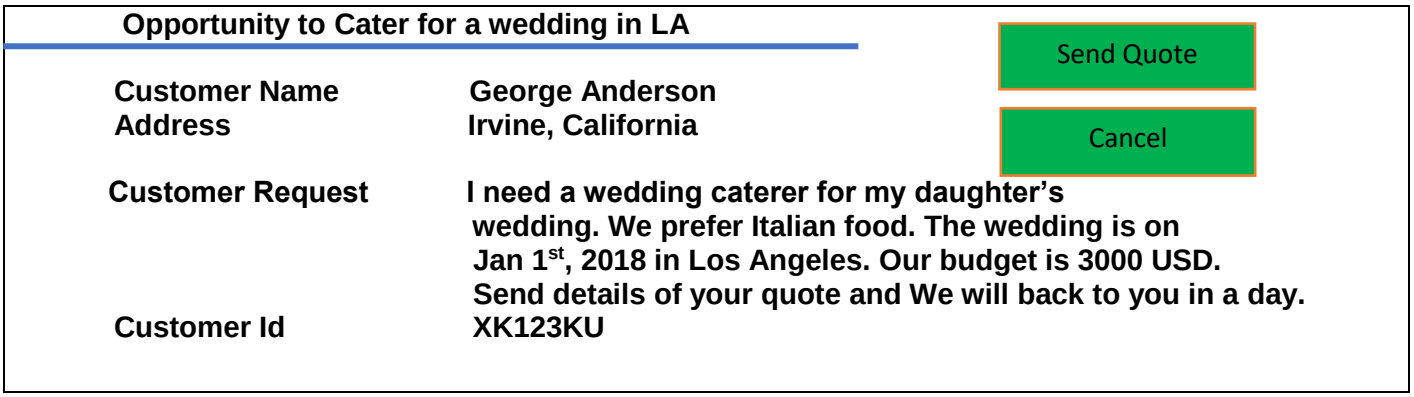
The button background color is changed to none for users.



Variation 2's observed conversion rate (2.89%) is 46.19% lower than the baseline conversion rate (5.38%). You can be 95% confident that this result is a consequence of the changes we made and not a result of random chance.

Variation 3:

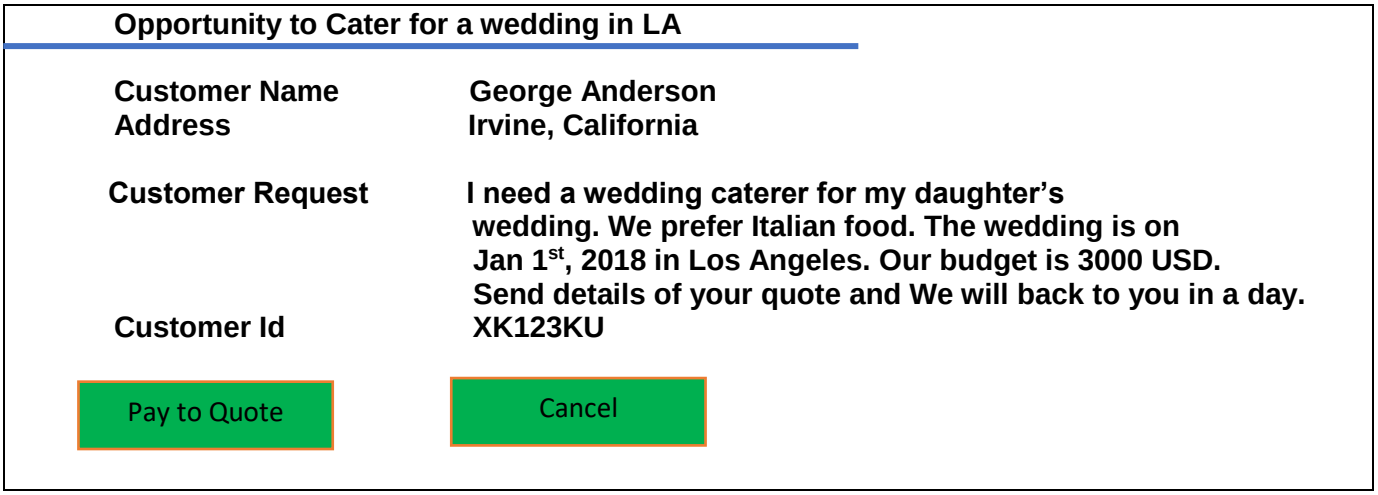
In this version, we change the button text to “Send Quote”.



Variation 3's observed conversion rate (8.42%) was 56.48% higher than the baseline conversion rate (5.38%). You can be 95% confident that this result is a consequence of the changes we made and not a result of random chance.

Variation 4:

The variation here will be the alignment of the "Pay to Quote" button at the bottom of the form.



The test result is not significant. The observed difference in conversion rate (22.24%) isn't big enough to declare a significant winner. There is no real difference in performance between the baseline and Variation 4.

From the above results and metric, the variation 3 is the winner among the 4 versions of the design. But, the conversion rate can be increased furthermore by testing more variations. In the future variation, the Variation 3 can be made the baseline and changes can be tested against it.

If I were to run the test again, I would

- Do a survey on the things customers do not like about the form
- Change the main button color to blue

- Run it on 1000 users instead of 606 users as in Variation 3 which is the new baseline. By increasing the sample size, we reduce the error and hence can have greater confidence that our sample result is going to be close to the true mean.

Do you have any thoughts on the experimental design?

When running a testing programme we will have to decide for each test whether it should be run as an A/B test or a multivariate test. The decisions made when designing our experiments will significantly impact important variables such as the

1. depth of the insights,
2. the speed of testing and
3. delivering winning variations
4. the impact of your testing efforts.

A/B tests can run in a few different ways either as single tests on single features, run as a series of tests (e.g. A/B, B/C, etc.) or run as straight A/B/C/n tests.

The test used in my analysis is **A/B/C/D/E** test.

Multivariate testing allows us to combine multiple variations for multiple sections of the page and identify which combination of variations combines to deliver greatest impact.

Full-factorial multivariate testing is the most common form which means if you have 3 variations in 3 sections then you will have 27 variations. Each variation will receive one-twenty-seventh of the test traffic.