

Design of Experiments for Digital Marketing Using JMP®

Anish Koppula, Starcom MediaVest Group, Chicago, IL

Erich Gundlach, SAS Institute Inc., Cary, NC

ABSTRACT

Creative testing is an integral part of every digital marketing strategy. Evaluation of the performance of media assets is vital so that media dollars can be effectively allocated across the test groups. Today, most marketers rely on simple A/B tests to compare the performance of ad-copies using traditional hypothesis testing. A design of experiments approach provides a more efficient and informative comparison of ad-copies, thus allowing advertisers to effectively execute creative rotation and measure consumer response. This idea can also be extended to web analytics, which requires optimization of a set of elements on a website to provide the best landing page experience, ultimately leading to improved conversions and return on ad investment. This paper proposes and provides examples of: (1) various experimental designs (DOEs) replacing A/B tests, which report on efficiency in execution times using minimum sample test size determinations; and (2) methods for analyzing the DOE data that allow better visualization of the results and more informed decision making compared to methods used in A/B testing. JMP is used at both the design and analysis phases of each example to report on statistical measures aiding decision making, including sample size, power and correlations, along with other practical considerations.

INTRODUCTION

Digital marketers deploy multiple variations of the advertising units on the web to engage potential customers by driving them to the brand pages and thus facilitating the purchase of their products. It is often observed that the ad unit with a good appealing message and a strong call to action tends to bring in more conversions. In this regard, it is vital that marketers measure the effectiveness of these ad units in order to allocate spend accordingly and maximize the return on ad investment. During the period live in the market, guessing the result of the test by quickly comparing the performance of the variations has been a common mistake. The use of robust sample design practices allows more efficient use of time and money on experiments. This will also help us to arrive at a more reasonable conclusion on a winner by investing optimally.

The marketer's view towards measurement can always get shortsighted due to the pressure to jump to conclusions based on differences that are optimal but not real which means that results are probably due to random variation. This need of swift results in the testing environments is driven by budget constraints and the urgency to take business recommendations to the advertisers. The change in KPIs is also an often consideration in measuring the effectiveness of the copies. The ultimate goal is to obtain reliable results with statistical significance in a quick span of time while refraining to guess the winner by just eyeballing the average performance of the variations. Traditionally, A/B tests have been used to test the difference in the performance of two ad copies while a one-way ANOVA helped to measure the same for more than two variations. These tests lack the ability to report on the contributing factor in a winning combination and usually involve iterative execution of the test factors in combining sequences which might result in longer time and significant investments before a reliable decision on the test could be made. The concept of DOE has proved to be a more effective approach to address these challenges.

This paper illustrates specific use cases to address the minimum sample size determinations and the hypothesis tests needed to determine the significant difference between the performances of the factors. The examples are demonstrated using analytical workflows in JMP and the results of the tests are transformed into required recommendations and business insights. The change in the results can be observed depending on increase in the sample size and varying the power of the test. The design of experiments in each of the scenarios is the optimal way of setting up the test for reliable results with multiple responses in the measurement framework.

EXPERIMENTAL SETUP AND DATA

The experimental set up described in this paper is a hypothetical marketing campaign launched for a SAS Analytics conference. It lays out the advertising efforts on various digital channels with conversion and awareness strategies running on three different sites. Variations of ad copies with different colors, messaging and call to actions are launched to test the effectiveness of a copy among the available options.

Conversion strategy aims at bringing in online registrations from the customer paths while the awareness strategy is intended to make the potential attendees aware of the conference information and let them engage with the website. The three partners on which the advertising is carried out are identified as AllAnalytics, KDNuggets and Google. Online display banners are allocated to AllAnalytics and KDNuggets in conjunction with a strong SEM marketing presence on Google. The creative copies cater to three messages on “Early Bird Registration”, “Come Join Us” and “Network with the Analytics Community”. The display banners are deployed in three placements corresponding to sizes 160X600, 300X250 and 728X90. The two call-to-actions either ask the audience to “Register Now” or to “Learn More”. The two colors intended to test with the logos are blue and orange.

The performance of these media assets is primarily evaluated on the basis of Click Through Rate (CTR) as a Key Performance Indicator (KPI), which is calculated as the number of clicks obtained to the total number of impressions delivered. The secondary metric used for evaluation is the Conversion Rate (CR), which is calculated as the number of registrations recorded in the pool of impressions served to the audience. The difference of the performance is measured after collecting a reliable sample size.

It is assumed that the business would be interested in observing a difference of 0.04% or higher in order to treat the performance of two variations different. The impressions are assumed to be normally distributed and the noise in the experiment is considered random. It is also assumed that all other factors in the ecosystem remain unchanged while measuring the impact of media assets. Seasonality and other external factors that might bias the performance of the test groups is not a part of this study and should be addressed through a separate time series analysis as desired

SAMPLE SIZE ANALYSIS

As discussed earlier, the results of the test might vary significantly as we collect more samples in the experiment. It is vital that one determine the minimum sample size at the time of the launch to be confident of the time required to obtain sufficient data and avoid the risk of arriving at an incorrect conclusion.

In our hypothetical marketing campaign, with an observed historical CTR of 0.19% for creative A, Creative B is expected to bring in a click through rate of 0.23% or higher in order to be treated different. So the minimum number of impressions needed for the test would be obtained by plugging in these two values as desired proportions in the sample size calculations. Also note that the conclusions of the test might differ until the minimum sample size is reached, depending on time of measurement in the data collection process of the experiment. Clearly, we can observe that the earlier one attempts to draw conclusions, the greater is the risk of failing to detect the difference between the copies.

| Local Data Filter | | Contingency Analysis of Click Through By Creative | | | |
|----------------------------------|--|---|-------------------|--|--------------------|
| 100000 matching rows | | Tests | | | |
| <input type="checkbox"/> Inverse | | N | DF | -LogLike | RSquare (U) |
| 1 ≤ Row ≤ 100000 | | 100000 | 1 | 1.1589967 | 0.0008 |
| | | Test | ChiSquare | Prob>ChiSq | |
| | | Likelihood Ratio | 2.318 | 0.1279 | |
| | | Pearson | 2.313 | 0.1283 | |
| | | Fisher's | | | |
| | | Exact Test | Prob | Alternative Hypothesis | |
| | | Left | 0.9446 | Prob(Click Through=1) is greater for Creative=A than B | |
| | | Right | 0.0736 | Prob(Click Through=1) is greater for Creative=B than A | |
| | | 2-Tail | 0.1473 | Prob(Click Through=1) is different across Creative | |
| | | Two Sample Test for Proportions | | | |
| | | Proportion | | | |
| | | Description | Difference | Lower 95% | Upper 95% |
| | | P(11A)-P(11B) | -0.00042 | -0.00096 | 0.000124 |
| | | Adjusted | | | |
| | | Wald Test | | Prob | |
| | | P(11A)-P(11B) >= 0 | | 0.9349 | |
| | | P(11A)-P(11B) <= 0 | | 0.0651 | |
| | | P(11A)-P(11B) = 0 | | 0.1303 | |

Figure 1. Contingency analysis with filter applied to show results after 100,000 impressions.

| After N total impressions | $P(1 A)-P(1 B) \leq 0$ | $P(1 A)-P(1 B) = 0$ | $P(1 A)-P(1 B) \geq 0$ |
|---------------------------|------------------------|---------------------|------------------------|
| 5000 | 0.3693 | 0.7387 | 0.6307 |
| 10000 | 0.3907 | 0.7814 | 0.6093 |
| 100000 | 0.0651 | 0.1303 | 0.9349 |
| 200000 | 0.0207* | 0.0414* | 0.9793 |
| 680000 | 0.0004* | 0.0009* | 0.9996 |

Table 1. P-values after 5000; 10000; 100000; 200000; and 680000 impressions. Note that the correct conclusion [P(1|B) is significantly greater than P(1|A)] from this data set is not reached until approximately 200,000 impressions.

From the above illustration, we could learn that it is quite important that the marketers resist the tendency to make quick decisions. It is like a ball game where you let the match to finish and determine a winner. We can still keep track of the scores and get a sense of the result until a final judgment is made while the game is in progress.

The adverse effects of making early decisions can be avoided by determining an optimal sample size prior to going live. This will ensure confident measurement of a significant delta within the variations upon reaching that limit. Please note that higher the desired power for the test, the higher impressions we need to measure the test difference. It is also important to validate no change in the test scenario during the flight of the experiment to introduce a bias due to any additional factors outside of the test components.

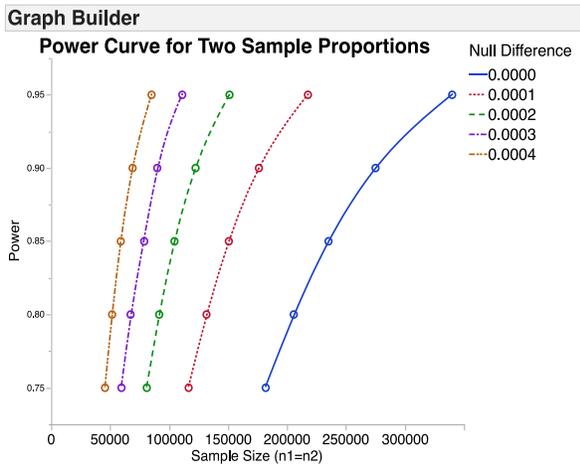


Figure 2

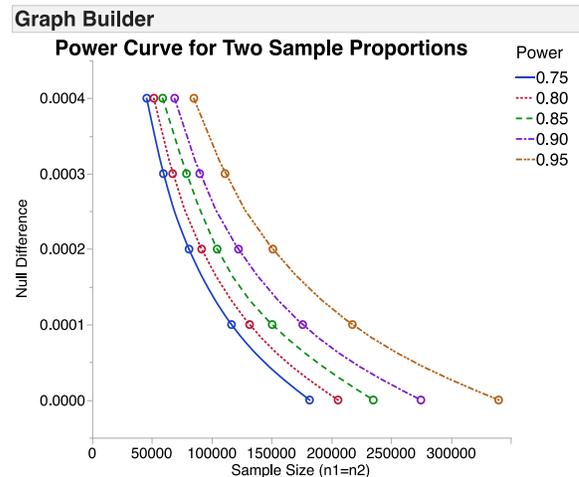


Figure 3

**Figure 2. Power curve showing power vs. sample relationship size for different null differences to detect
Figure 3. Power curve showing null difference vs. sample size relationship for desired power levels**

The power and sample size analysis in JMP helps us calculate the minimum sample size in the experiment in order to make a statistically significant conclusion in a hypothesis testing environment. In the given example, the two copies are tested for the difference in CTR with a minimum null difference of 0.04%. With desired power of 90 and using a historical click through rate benchmark of 0.19%, the minimum number of impressions required is 274,874.

Assuming that we could deliver 40k impressions per day, it would take about 14 days to collect sufficient samples. Please note that the minimum sample size could be reduced by compromising on the power of the test. Also, the smaller the null difference to test, the larger the required sample size.

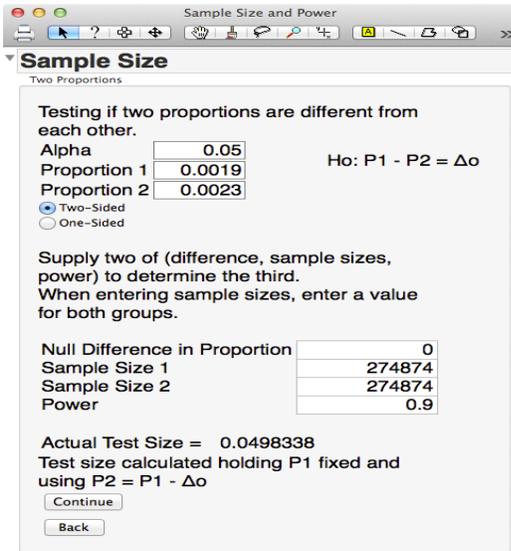


Figure 4. To determine the sample size, select **Sample Size and Power** from the DOE menu, then choose **Two Sample Proportions**. Fill out the dialog as above, leaving **Sample Size 1** and **Sample Size 2** blank. Click **Continue**.

The same calculations can be performed if the KPI was conversion rate instead of the click through rate. Since the CR values are much lower than the CTR values, we may need substantially more impressions to measure the difference in conversion rates.

The formula JMP uses to determine minimum sample size can be found in Clay (2014).

TRADITIONAL APPROACH VS. DESIGN OF EXPERIMENTS

Traditional testing focuses on reporting the better performing combination, while in a multivariate testing environment, we often are required not only to determine which combination is better combination, but also the measure the effect of each individual element. The general idea is to run two competing variations by changing one factor at a time, then run the next factor against the winner from the first run. This approach generally involves several iterations and takes a significant amount of time. Furthermore, it does not allow estimation of interactions between factors, possibly leading to erroneous conclusions.

DESIGN OF EXPERIMENTS

The concept of Design of Experiments can be illustrated using the DOE platform in JMP. The analytical workflow needed to measure the performance of the test can vary depending on the following considerations:

1. Time of launch of the variations – Two Challengers vs. Champion and Challenger
2. Number of factors changing at a time
3. Interactions due to change in more than one factor at a time

This paper discusses the findings from four example scenarios using clicks and impressions and/or CTR as the KPI:

- Two copies new in the market
- Three variations in a factor
- Two copies with the two factors varying at a time
- Multiple test variations deployed to measure the winner based on the best performing combination

TWO COPIES NEW IN MARKET

Scenario: It is usually hypothesized confidently that a strong call to action brings in more conversions. The team wants to prove or disprove this claim. A creative ad copy with the same color, messaging and strategy go live with two variations in the call to action: “Learn More” and “Register now.” They are executed with a 50-50 split which means that each of the variations is shown 50% of the time to the audience visiting the sites where they reside.

Objective: Which variation performs better?

Considerations: Benchmark Click Through Rate: 0.19%, Null Difference to Test: 0.04%

Test for significance: Contingency analysis or two sample t-test for proportions

The required sample size is 274,874 impressions for each copy. 275,000 impressions were measured for each copy, and the results are summarized, below.

| Creative | Click Through | N |
|----------|---------------|--------|
| A | 0 | 274716 |
| A | 1 | 284 |
| B | 0 | 274867 |
| B | 1 | 133 |

Table 2. Summary of results. Click Through is an indicator variable, where 0 indicates the viewer did not click and 1 indicates the viewer clicked. N is the number of impressions.

Creative A has the call to action “Learn More” while creative B has “Register Now.” After executing impressions greater than the required sample size, we could measure the performance by conducting relevant hypothesis tests.

The below graphic shows us the p-values for fisher’s exact test and adjusted wald’s test treating the Click Through indicator variable binary (0/1). This workflow assumes binomial distribution for the response. It was observed that the probability of click through being 1 is significantly greater for “Register Now” when compared to that “Learn More.”

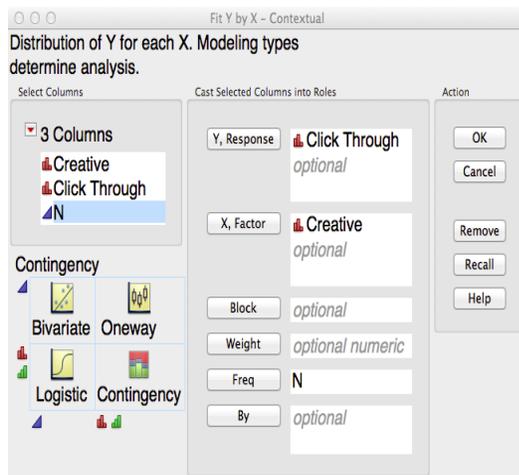


Figure 5. To perform a contingency analysis, select Fit Y by X from the Analyze menu, fill out the dialog as above, and click OK. Note that the modeling type for Click Through is set to Nominal.

Contingency Analysis of Click Through By Creative

Contingency Table

| | Click Through | | |
|------------|---------------|-------|--------|
| | 0 | 1 | |
| Count | | | |
| Total % | | | |
| Col % | | | |
| Row % | | | |
| Creative A | 339366 | 634 | 340000 |
| | 49.91 | 0.09 | 50.00 |
| | 50.01 | 45.55 | |
| | 99.81 | 0.19 | |
| Creative B | 339242 | 758 | 340000 |
| | 49.89 | 0.11 | 50.00 |
| | 49.99 | 54.45 | |
| | 99.78 | 0.22 | |
| | 678608 | 1392 | 680000 |
| | 99.80 | 0.20 | |

Tests

| N | DF | -LogLike | RSquare (U) |
|--------|----|-----------|-------------|
| 680000 | 1 | 5.5416453 | 0.0006 |

Test

| Test | ChiSquare | Prob>ChiSq |
|------------------|-----------|------------|
| Likelihood Ratio | 11.083 | 0.0009* |
| Pearson | 11.069 | 0.0009* |

Fisher's Exact Test

| Prob | Alternative Hypothesis |
|---------|--|
| 0.9996 | Prob(Click Through=1) is greater for Creative=A than B |
| 0.0005* | Prob(Click Through=1) is greater for Creative=B than A |
| 0.0010* | Prob(Click Through=1) is different across Creative |

Two Sample Test for Proportions

| Description | Difference | Lower 95% | Upper 95% |
|---------------|------------|-----------|-----------|
| P(11A)-P(11B) | -0.00036 | -0.00058 | -0.00015 |

Adjusted Wald Test

| P(11A)-P(11B) >= 0 | Prob |
|--------------------|---------|
| P(11A)-P(11B) <= 0 | 0.9996 |
| P(11A)-P(11B) = 0 | 0.0004* |
| P(11A)-P(11B) = 0 | 0.0009* |

Figure 6. After selecting the appropriate tests from the red triangle menu, it can be concluded that create B had a significantly higher proportion of clicks than creative A.

Now, if the data was collected on click through rate as a proportion instead of a binary variable representing the existence of a click or not. The testing needs to be replicated for a normally distributed continuous response. The below graphic shows us the p-values for a CTR (B) – CTR (A) t-test assuming equal variances. In this case, same conclusion was observed that the click through rate for “Register Now” is significantly greater than that of the copy with a “Learn More” call to action.

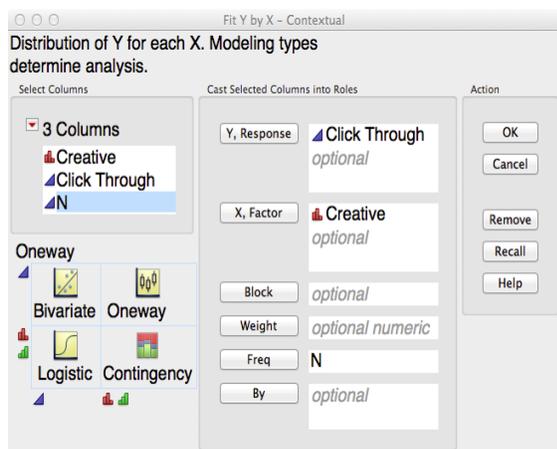


Figure 7. To perform a t Test, select Fit Y by X from the Analyze menu, fill out the dialog as above, and click OK. Note that the modeling type for Click Through is set to Continuous.

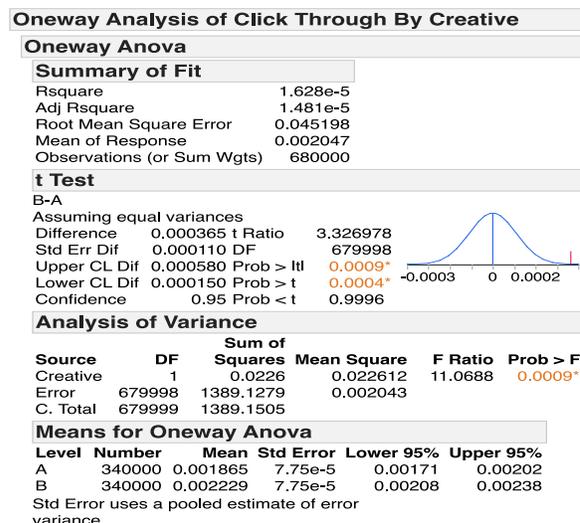


Figure 8. After selecting the appropriate tests from the red triangle menu (in this case a t Test), it can be concluded that create B had a significantly higher proportion of clicks than creative A.

It can be noted that the t-test not only arrives at the p-value but also reports the upper and lower limits based on the confidence interval.

THREE VARIATIONS IN A FACTOR

Scenario: The marketing team now wants to know if the size of the copy influences the customer interaction. To execute this, the creative copy with unchanged color and messaging but with the “Register Now” winner call to action goes live in three placements on the sites with sizes of 160x600, 300x250, and 728x90. They are executed in a 1:1:1 ratio, with each variation shown with equal probability to the audience visiting the sites where they reside.

Objective: Which variation performs better?

Considerations: Benchmark Click Through Rate: 0.19%, Null Difference to Test: 0.04%

Test for significance: One Way Analysis of Variance (ANOVA)

| Placement | Click Through | N |
|-----------|---------------|--------|
| 160x600 | 0 | 274326 |
| 160x600 | 1 | 674 |
| 300x250 | 0 | 274546 |
| 300x250 | 1 | 454 |
| 728x90 | 0 | 274701 |
| 728x90 | 1 | 400 |

Table 3. Summarized results. Note: Sample size analysis was performed as in the previous example.

In this case, it should be noted that we save significant time and impressions by running one ANOVA test instead of two sequential t-tests in order to measure the significance between three variations in a factor. With 275000 impressions simultaneously run against each of the variations (minimum sample size of 274874), we are now confident to measure the difference between their performance using ANOVA.

The below graphic shows us the LSD Threshold matrix and the connecting letters report from JMP. Both the representations are used to depict differences in pairwise comparisons using Student's t distribution in one way analysis of variance. JMP offers additional examples to interpret ANOVA output using the Oneway platform. Clearly, in this scenario, we could conclude that the 160X600 has a significantly better click through rate when compared to the other two variations. The mean values could also be observed in the connecting letters report.

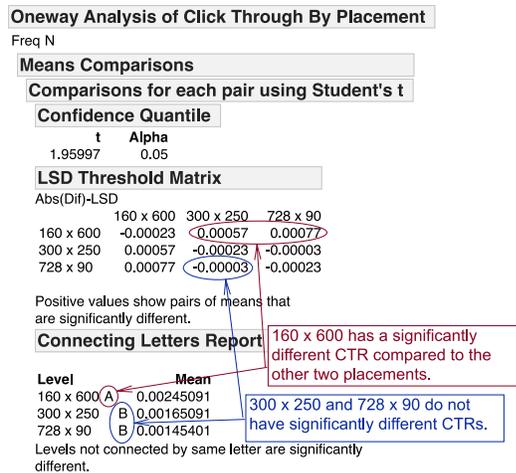


Figure 9. Perform a Oneway Analysis, as described previously, using Fit Y by X (after setting Click Through to modeling type of Continuous) to calculate the results.

TWO COPIES WITH TWO FACTORS VARYING AT A TIME

Scenario: Based on the results from above two scenarios, the marketing team wants to measure the true impact of a factor in the copy knowing that color has been an influencing factor from a past analysis. This time, they want to test four variations, two each with a varying color and call to action. They tested orange color with Register Now/ Learn More and then blue color with Register Now/ Learn More

Objective: Which combination performs better? Is color or call to action the contributing factor in the winner?

Considerations: Other factors, like placement and site, remain the same across the experiment

Test for significance: GLM or Two Way ANOVA followed by back tests

| Color | Call to Action | Clicks | Impressions |
|--------|----------------|--------|-------------|
| blue | Register Now | 203 | 62100 |
| orange | Register Now | 166 | 62100 |
| orange | Learn More | 69 | 62100 |
| blue | Learn More | 83 | 62100 |

Table 4. 62,100 impressions were made of each creative, and the clicks were recorded.

The minimum sample size required for this test scenario is calculated using the formula proposed in reference paper 'Sample size estimates for 2^{k-p} designs with binary responses'. We determined that the measurement in performance across these four variations could be reliable only after executing approximately 62k impressions per each of the creative copy combinations. Once we achieved that, we followed few simple steps using JMP's DOE framework in order to design an experiment to meet the specifications in the scenario.

1. Select Custom Design from JMP's DOE menu and fill out the dialog as below.
2. Note that the interaction between the factors was included in the model.
3. Choose the minimum of 4 runs, since there are only 4 possible combinations of the two factors.
4. Click Make Design.

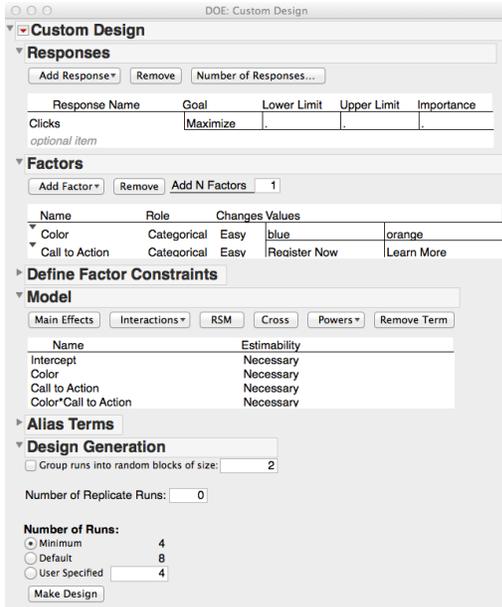


Figure 11. Dialog to make custom design in JMP's DOE Menu.

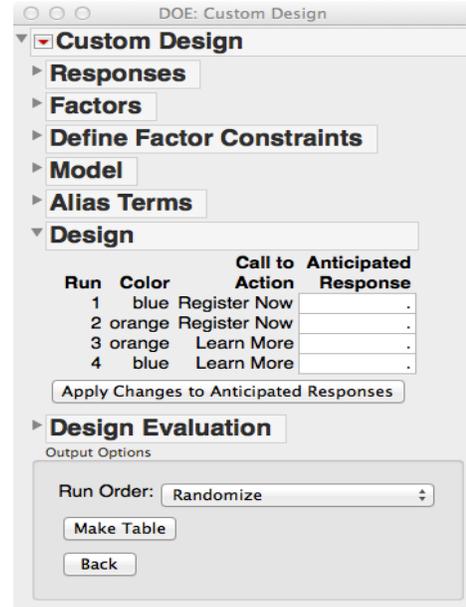


Figure 10. The resulting design.

After making the design, we need to use the factors in the scenario to build a model for a reasonable fit to the data and thus determine significant effects. For illustration purposes, a generalized linear model is chosen here. It includes color, call to action as main effects and a two degree interaction effect against clicks as a response.

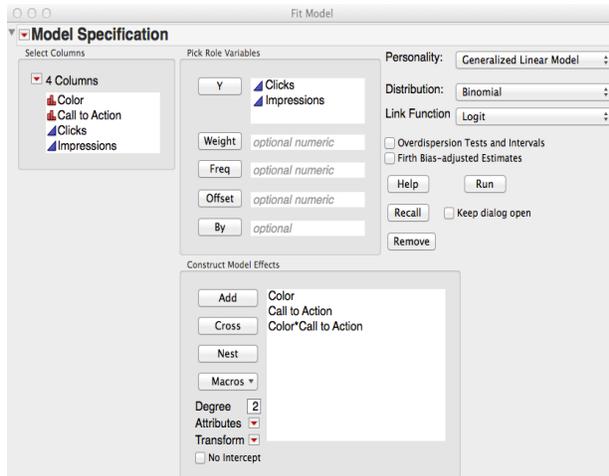


Figure 12. To model the data, select Fit Model from JMP's Analyze menu. Fill out the dialog as above.

Generalized Linear Model Fit

Response: Clicks
 Distribution: Binomial
 Link: Logit
 Estimation Method: Maximum Likelihood
 Observations (or Sum Wgts) = 4

Whole Model Test

| Model | -LogLikelihood | ChiSquare | DF | Prob>ChiSq |
|------------|----------------|-----------|----|------------|
| Difference | 49.2027978 | 98.4056 | 3 | <.0001* |
| Full | 13.2137488 | | | |
| Reduced | 62.4165466 | | | |

Goodness Of Fit

| Fit Statistic | ChiSquare | DF | Prob>ChiSq |
|---------------|-----------|----|------------|
| Pearson | 0.0000 | 0 | . |
| Deviance | 0.0000 | 0 | . |

AICc

Effect Tests

| Source | DF | ChiSquare | Prob>ChiSq |
|----------------------|----|-----------|------------|
| Color | 1 | 4.0020173 | 0.0454* |
| Call to Action | 1 | 92.376937 | <.0001* |
| Color*Call to Action | 1 | 0.0075638 | 0.9307 |

Parameter Estimates

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq | Lower CL | Upper CL |
|--|-----------|-----------|-----------|------------|-----------|-----------|
| Intercept | -6.284866 | 0.0484482 | 336986.15 | <.0001* | -6.361641 | -6.171638 |
| Color[blue] | 0.0986939 | 0.0484482 | 4.0020173 | 0.0454* | 0.0019572 | 0.1920878 |
| Call to Action[Register Now] | 0.4439371 | 0.0484482 | 92.376937 | <.0001* | 0.3501599 | 0.5402283 |
| Color[blue]*Call to Action[Register Now] | 0.004214 | 0.0484482 | 0.0075638 | 0.9307 | -0.091079 | 0.090539 |

Figure 13. Results from the GLM model fit, indicating that "blue" color and "Register Now" call to action are significant while their interaction is not.

The generalized linear model reveals that the blue variation in the color factor and the register now variation in the call to action factor are significant independently. However, the ad copy with “blue” color and a “register now” call to action does not perform better. So, executing a change in the color and the call to action in an ad copy at the same will not potentially improve the clicks.

Knowing that the interaction effect is not significant, we could refine the model fit by removing the interaction and just including the main effects to recalculate the p-values and then choose the estimates. Further, the prediction profiler in JMP allows us to simulate the responses based on different variations in predictors or independent variables in the model. We can observe that the probability of a click is 0.32% when the color is “blue” and the call to action is “Register Now.” The response value for another variation could be obtained by using the sliders for each of the effects in the prediction profiler.

Generalized Linear Model Fit

Response: Clicks
Distribution: Binomial
Link: Logit
Estimation Method: Maximum Likelihood
Observations (or Sum Wgts) = 4

Whole Model Test

| Model | -LogLikelihood | ChiSquare | L-R | DF | Prob>ChiSq |
|------------|----------------|-----------|-----|----|------------|
| Difference | 49.199016 | 98.3980 | | 2 | <.0001* |
| Full | 13.2178306 | | | | |
| Reduced | 62.4165466 | | | | |

Goodness Of Fit

| Fit Statistic | ChiSquare | DF | Prob>ChiSq |
|---------------|-----------|----|------------|
| Pearson | 0.0076 | 1 | 0.9307 |
| Deviance | 0.0076 | 1 | 0.9307 |

AICc

Effect Tests

| Source | DF | ChiSquare | Prob>ChiSq |
|----------------|----|-----------|------------|
| Color | 1 | 5.0126439 | 0.0252* |
| Call to Action | 1 | 93.387213 | <.0001* |

Parameter Estimates

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq | Lower CL | Upper CL |
|------------------------------|-----------|-----------|-----------|------------|-----------|-----------|
| Intercept | -6.265028 | 0.0484238 | 336966.15 | <.0001* | -6.361737 | -6.171835 |
| Color[blue] | 0.098447 | 0.044076 | 5.0126439 | 0.0252* | 0.0122463 | 0.185128 |
| Call to Action[Register Now] | 0.4443419 | 0.0482319 | 93.387213 | <.0001* | 0.3509724 | 0.5401863 |

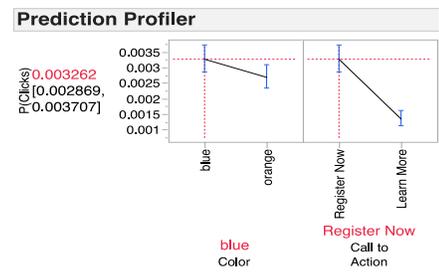
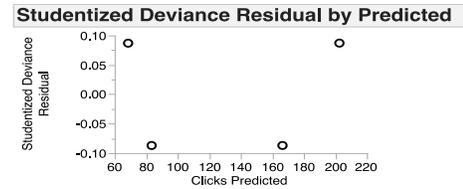


Figure 14(a, b). To obtain the refined model, repeat the analysis, as above, this time leaving out the interaction.

CUSTOM DESIGN ILLUSTRATING A WINNER

Scenario: At last, the SAS conference marketing team wants to test the most ideal combination across all the factors in the experiment. Deploying the variations of the copies with available colors, placements, sites, call to actions and messages, and the intention is to test the combination of the factors that work best

Objective: What is the best possible combination of the factors?

Considerations: Prior performance benchmarks for each of the individual combinations partially known

Test for significance: Custom design with GLM

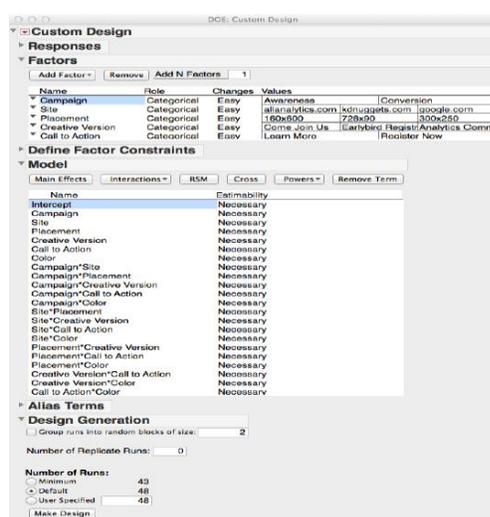


Figure 15. Dialog to make custom design in JMP's DOE Menu

| Run | Campaign | Site | Placement | Creative Version | Call to Action | Color | Anticipated Response |
|-----|------------|------------------|-----------|------------------------|----------------|--------|----------------------|
| 1 | Awareness | google.com | 728x90 | Come Join Us | Register Now | Orange | -2 |
| 2 | Conversion | kduggets.com | 300x250 | Earlybird Registration | Learn More | Blue | -2 |
| 3 | Awareness | allanalytics.com | 160x500 | Analytics Community | Learn More | Blue | -2 |
| 4 | Conversion | google.com | 160x500 | Analytics Community | Register Now | Orange | -2 |
| 5 | Awareness | google.com | 300x250 | Analytics Community | Register Now | Blue | 1 |
| 6 | Conversion | allanalytics.com | 728x90 | Earlybird Registration | Learn More | Orange | 0 |
| 7 | Awareness | allanalytics.com | 728x90 | Earlybird Registration | Register Now | Orange | 0 |
| 8 | Conversion | kduggets.com | 160x500 | Analytics Community | Register Now | Blue | -9 |
| 9 | Conversion | google.com | 300x250 | Analytics Community | Learn More | Orange | -1 |
| 10 | Awareness | google.com | 160x500 | Come Join Us | Register Now | Blue | 6 |
| 11 | Awareness | allanalytics.com | 728x90 | Analytics Community | Register Now | Blue | -2 |
| 12 | Awareness | kduggets.com | 300x250 | Earlybird Registration | Register Now | Blue | -2 |
| 13 | Conversion | google.com | 728x90 | Earlybird Registration | Learn More | Orange | -2 |
| 14 | Awareness | google.com | 300x250 | Come Join Us | Learn More | Orange | -3 |
| 15 | Conversion | allanalytics.com | 160x500 | Analytics Community | Register Now | Orange | -6 |
| 16 | Awareness | allanalytics.com | 300x250 | Earlybird Registration | Learn More | Blue | 6 |
| 17 | Awareness | google.com | 160x500 | Analytics Community | Register Now | Orange | -3 |
| 18 | Awareness | kduggets.com | 728x90 | Analytics Community | Register Now | Orange | -2 |
| 19 | Awareness | kduggets.com | 728x90 | Come Join Us | Register Now | Blue | -4 |
| 20 | Conversion | kduggets.com | 300x250 | Come Join Us | Register Now | Orange | 2 |
| 21 | Conversion | google.com | 160x500 | Earlybird Registration | Register Now | Blue | 0 |
| 22 | Awareness | allanalytics.com | 500x250 | Come Join Us | Register Now | Blue | 4 |
| 23 | Conversion | google.com | 728x90 | Analytics Community | Register Now | Blue | 1 |
| 24 | Awareness | kduggets.com | 160x500 | Earlybird Registration | Learn More | Blue | 2 |
| 25 | Conversion | allanalytics.com | 300x250 | Analytics Community | Learn More | Orange | -2 |
| 26 | Conversion | google.com | 160x500 | Come Join Us | Register Now | Blue | -1 |
| 27 | Conversion | allanalytics.com | 160x500 | Earlybird Registration | Learn More | Blue | 6 |
| 28 | Awareness | allanalytics.com | 300x250 | Analytics Community | Register Now | Orange | -5 |
| 29 | Awareness | google.com | 728x90 | Earlybird Registration | Learn More | Blue | -2 |
| 30 | Conversion | google.com | 160x500 | Come Join Us | Register Now | Blue | 0 |
| 31 | Conversion | google.com | 728x90 | Come Join Us | Learn More | Blue | 0 |
| 32 | Conversion | allanalytics.com | 300x250 | Earlybird Registration | Register Now | Orange | 0 |
| 33 | Conversion | kduggets.com | 160x500 | Come Join Us | Learn More | Blue | 2 |
| 34 | Conversion | allanalytics.com | 300x250 | Come Join Us | Learn More | Blue | 2 |
| 35 | Conversion | google.com | 160x500 | Analytics Community | Learn More | Blue | 1 |
| 36 | Awareness | kduggets.com | 728x90 | Earlybird Registration | Learn More | Blue | -8 |
| 37 | Awareness | kduggets.com | 160x500 | Come Join Us | Register Now | Orange | -2 |
| 38 | Awareness | google.com | 160x500 | Earlybird Registration | Learn More | Orange | 2 |
| 39 | Awareness | allanalytics.com | 728x90 | Analytics Community | Register Now | Orange | 1 |
| 40 | Awareness | allanalytics.com | 160x500 | Come Join Us | Learn More | Orange | 10 |
| 41 | Conversion | allanalytics.com | 728x90 | Analytics Community | Learn More | Blue | 2 |
| 42 | Conversion | allanalytics.com | 728x90 | Come Join Us | Register Now | Orange | 0 |
| 43 | Conversion | kduggets.com | 728x90 | Earlybird Registration | Register Now | Blue | -4 |
| 44 | Awareness | kduggets.com | 300x250 | Earlybird Registration | Register Now | Orange | -1 |
| 45 | Conversion | allanalytics.com | 300x250 | Analytics Community | Register Now | Blue | -5 |
| 46 | Conversion | kduggets.com | 728x90 | Come Join Us | Learn More | Orange | 3 |
| 47 | Awareness | kduggets.com | 300x250 | Analytics Community | Learn More | Orange | 3 |
| 48 | Awareness | allanalytics.com | 728x90 | Come Join Us | Learn More | Blue | 12 |

Figure 16. The resulting design.

To accomplish the above objective, we need to create a custom design similar to the third scenario but now using all the factors and their interactions in the design with a default of 48 runs. Once the design is made, we would then build a model that would best fit the data. In the first iteration, we used a generalized linear model with binomial distribution and a logit link function to force fit all the main and interaction effects into the model to check the significance for each of them.

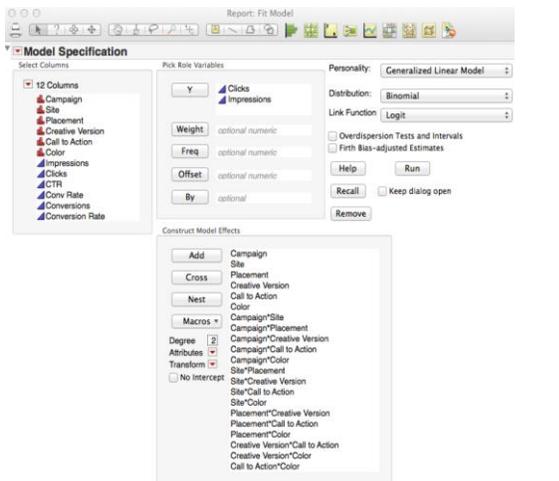


Figure 17. To model the data, select Fit Model from JMP's Analyze menu. Fill out the dialog as above.

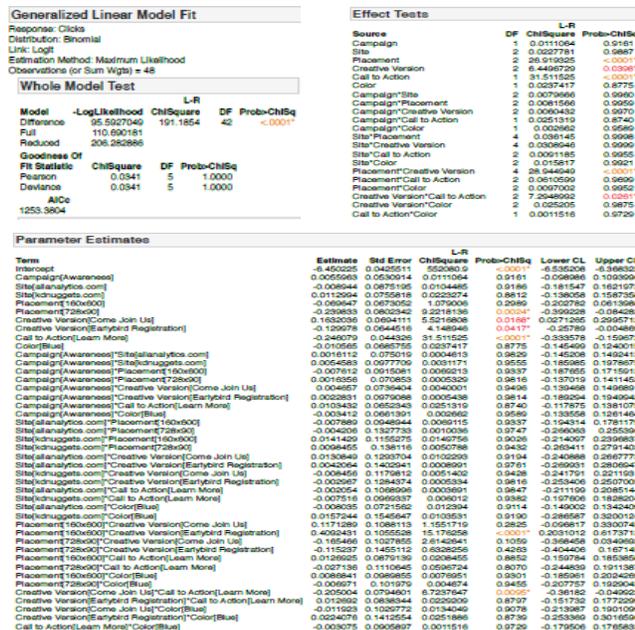


Figure 18. Results from the GLM model fit, indicating the significant and insignificant effects

The model could be simplified by including only significant effects from the first iteration and the p-values could be recalculated. The final effects that seem significant in the model are the main effects for placement, creative version and call to action along with the interactions between placement*creative and placement*call to action. "Come Join Us" creative messaging seems to work well with a "Learn More" call to action. The placement that brings optimal performance with this combination is identified as 728X90.

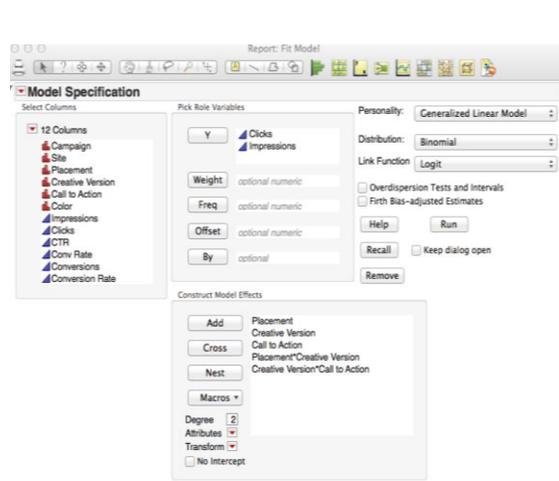


Figure 19. Refit the model with the significant main and interaction effects as shown in the dialog above.

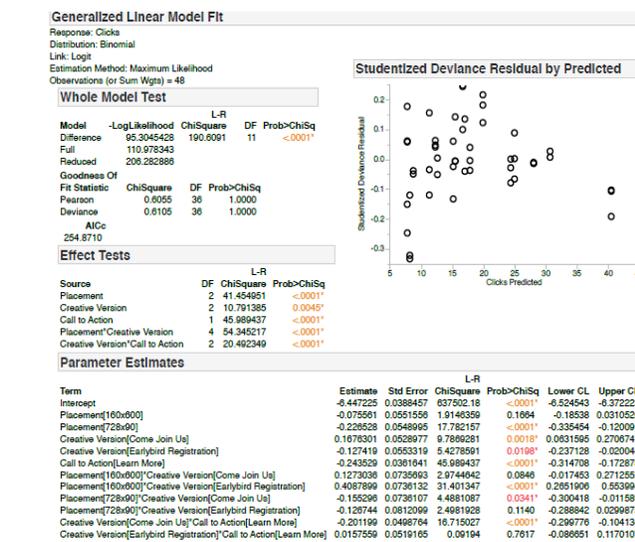


Figure 20. Results from the GLM model fit, indicating the final model effects.

CONCLUSION

This paper is a guide to demonstrate a workflow to efficiently measure the effectiveness of the marketing campaigns involving variations in ad copies. The design of experiments framework is put forth taking into consideration the time to market and the desired power to carry out the analysis. Actions and recommendations upon the execution of the ad copies are constructed in the form of business insights from the test results. Each of the discussed scenarios demonstrates a unique situation to measure the performance in variations.

REFERENCES

1. Barker, Clay, "Power and Sample Size Calculations in JMP," SAS Institute. Retrieved July 27, 2014 from http://www.jmp.com/blind/whitepapers/wp_jmp_powersample_104887.pdf.
2. Bisgaard, S. and H. T. Fuller (1995). Sample size estimates for 2^{k-p} designs with binary responses. *Journal of Quality Technology*, **27**(4), 344-354.

RECOMMENDED READING

Ledolter, Johannes and Swersey Arthur J. (2007), *Testing 1-2-3: Experimental Design with Applications in Marketing and Service Operations*, Stanford University Press: Stanford, CA.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Anish Koppula
Orbitz Worldwide, Inc.
(312) 894-2943
anish.koppula@orbitz.com

Erich Gundlach
SAS Institute Inc.
(919) 531-9412
erich.gundlach@jmp.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.