

1. What questions would you ask me about my goals and methodology?

- Is sampling for each variation random?
- What confidence level do you want to use in order to draw conclusion? 95% is the most common one to use.
- Have you thought about segmenting the viewers in order to identify segment-specific form tuning?
- Why are the sample sizes unequal?

2. Do you have any thoughts on the experimental design?

- 1) Make sure the sampling is random, and divided equally
- 2) I would use a bigger sample size if possible
- 3) To double check if the test result is reliable, I would test two identical variations using different sample data sets. If there is a large difference between the results, the results should not be trusted
- 4) I would let the test run for a while (instead of a one-time test) so that I can get test results on different samples

3. Interpretation of results and conclusion?

- **Assumption and terms:**

Random sampling for each variation

Each viewer makes independent quote decision

We want to use 95% as the confidence thresholds

“Population mean”: the baseline quote rate

“Sample mean”: the variation form quote rate

- **Interpretation and conclusion:**

A split test (or hypothesis testing) is conducted within four variations and the baseline.

Null hypothesis statement:

The changes in a variation form will **not** affect requests' quote rate

Alternative hypothesis statement:

The changes in a variation form will affect requests' quote rate

Alpha = 0.05

Baseline: 32 conversions out of 595 viewers

Population mean = 5.38%

Variation 1: 30 conversions out of 599 viewers

Sample mean = 5.01%

p-value = 0.386699

The probability value (P-value) is 0.386699, which is a lot higher than the chosen alpha (0.05). it means that probability of observing the sample distribution of Variation 1 in the baseline distribution is very high, which indicates that the sample mean is from the same distribution (baseline distribution). We got different mean only due to random variability (chance). In other words, the sample data from Variation 1 is **not** statistically significant judging by 95%

confidence level. Therefore, we **cannot** reject the null hypothesis statement: The changes in a variation form will not affect requests' quote rate.

Variation 2: 18 conversions out of 622 viewers

Sample mean = 2.89%

p-value = 0.014892

The probability value (P-value) is 0.014892, which is lower than the chosen alpha (0.05). It means that the sample mean might have come from other distribution. The other distribution might be anything other than baseline distribution. Therefore, in the case of Variation 2, we can confidently reject the null hypothesis, and therefore we believe that the changes Variation 2 form will affect requests' quote rate. Because the z score is negative, it means that the other distribution is located lower than the baseline mean, i.e. Variation 2 is negatively affecting the quote rate.

Variation 3: 51 conversions out of 606 viewers

Sample mean = 8.42%

p-value = 0.018636

The probability value (P-value) is 0.018636, which is lower than the chosen alpha (0.05). It means that the sample mean might have come from other distribution. The other distribution might be anything other than baseline distribution. Therefore, in the case of Variation 3, we can confidently reject the null hypothesis, and therefore we believe that the changes Variation 3 form will affect requests' quote rate. Because the z score is positive, it means that the other distribution is located higher than the baseline mean, i.e. Variation 3 is positively affecting the quote rate.

Variation 4: 38 conversions out of 578 viewers

Sample mean = 6.57%

p-value = 0.193855

The probability value (P-value) is 0.193855, which is higher than the chosen alpha (0.05). it means that probability of observing the sample distribution of Variation 4 in the baseline distribution is high, which indicates that the sample mean is from the same distribution (baseline distribution). We got different mean only due to random variability (chance). In other words, the sample data from Variation 4 is **not** statistically significant judging by 95% confidence level. Therefore, we **cannot** reject the null hypothesis statement: The changes in a variation form will not affect requests' quote rate.

4. Please provide statistical justification for your conclusions and explain the choices you made in your analysis.

1) Because the sample size is large enough compare to the population size in this test, we can apply Central Limit Theorem to the data. This means the *sample means* shape a normal distribution with the *population mean* as the mean.

2) Since Central Limit Theorem can be applied, we can use the calculation below to come up with the z-scores and p-values for each variation.

3) Calculation:

Experiment	trials	success
Baseline	595	32
Variation 1	599	30
Variation 2	622	18
Variation 3	606	51
Variation 4	578	38

Experiment	conversion rate	Standard error	p-value	z score
Baseline	5.38%	0.92%	-	-
Variation 1	5.01%	0.89%	0.386699	-0.2879339
Variation 2	2.89%	0.67%	0.014892	-2.1729399
Variation 3	8.42%	1.13%	0.018636	2.08278213
Variation 4	6.57%	1.03%	0.193855	0.86377832

4) Formulas used:

- Calculate standard errors:
 - $SE = \sqrt{p(1-p)/n}$
- p = conversion rate
- n = sample size
- Calculate z-score:
 - $z\text{-score} = \frac{\text{variation_p} - \text{baseline_p}}{\sqrt{\text{POWER}(\text{baseline_se}, 2) + \text{POWER}(\text{variation_se}, 2)}}$
- Calculate p-value:
 - $p\text{-value} = \text{NORMDIST}(z_score, 0, 1, \text{TRUE})$