Assuming all things equal, meaning the data variations only came from the quote form, then we can come to the following conclusion:

Bucket	Number of Quotes	Total number of viewers	Conversion Rate	Conversion Rate %
Baseline	32	595	0.05378151261	5.378151261
Variation 1	30	599	0.05008347245	5.008347245
Variation 2	18	622	0.02893890675	2.893890675
Variation 3	51	606	0.08415841584	8.415841584
Variation 4	38	578	0.06574394464	6.574394464

We are interested in answering two questions. Did the variation of quotes occur from design differences or strictly by chance? How large is the variation if it did not occur by chance? In statistics, there are no guarantees, so we have to change the word "strictly" to "unlikely" in the first question. Statisticians came to the consensus that an event is unlikely if it occurs less than 5% of the time. We ask the second question because everything in the real world will naturally vary in small amounts. Hence, we only care about large variations that are not expected. We can combine the first two questions to test for statistical significance, which means the variations of the designs were unlikely to have occurred by chance if the variations were expectantly small. A binomial test will be used since we are comparing two different proportions. For instance, the ratio between the number of quotes and total number of views differ for each variation. In addition, we are going to test with 95% confidence for the reason mentioned above. The p value represents the probability that the variations occurred by chance when the variations are expectantly small. Hence, a p value below .05 is statistically significant. Additionally, the ratio of quotes per views vary positively and negatively so a two tailed hypothesis test needs to be used. The R function prop.test() is a function used for the binomial test and will be used to run our analysis shown on the second page.

Variation 2 and Variation 3 to have a p value of `0.04153` and `0.04983` respectively. This means that 4% of the time, variation 2 and variation 3 will occur by chance, when variations are at its expected levels. Since those numbers are below .05, it is unlikely that these variations were due to chance. Hence, we can conclude that variation 2 will generate less quotes, and variation 3 will generate more quotes 95% of the time.

There are numerous valid questions to challenge our overly simplified assumption:

Did providers of different quote forms have the same economic incentives?

Were the same type of requests being made at the same price?

Were the invites selected randomly? If so, what method of random sampling was used?

How were the 3000 requests divided among the providers?

My thoughts are that the experiment should be designed in a way that minimizes all forms of variation other than the different designs of the quote form.

```
 prop.test(x = c(32,30), n =c(595,599))

        2-sample test for equality of proportions with continuity
        correction

data:  c(32, 30) out of c(595, 599)
X-squared = 0.024814, df = 1, p-value = 0.8748
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.02314956  0.03054564
sample estimates:
    prop 1     prop 2
0.05378151 0.05008347

> prop.test(x = c(32,18), n =c(595,622))

        2-sample test for equality of proportions with continuity
        correction

data:  c(32, 18) out of c(595, 622)
X-squared = 4.1541, df = 1, p-value = 0.04153
alternative hypothesis: two.sided
95 percent confidence interval:
 0.0007906984 0.0488945133
sample estimates:
    prop 1     prop 2
0.05378151 0.02893891

> prop.test(x = c(32,51), n =c(595,606))

        2-sample test for equality of proportions with continuity
        correction

data:  c(32, 51) out of c(595, 606)
X-squared = 3.847, df = 1, p-value = 0.04983
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.060627950 -0.000125856
sample estimates:
    prop 1     prop 2
0.05378151 0.08415842

> prop.test(x = c(32,38), n =c(595,578))

        2-sample test for equality of proportions with continuity
        correction

data:  c(32, 38) out of c(595, 578)
X-squared = 0.54968, df = 1, p-value = 0.4584
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.04081128  0.01688642
sample estimates:
    prop 1     prop 2
0.05378151 0.06574394


>
```