1 newpage

1.1 Part 11: The A/B Test (Rocket Fuel case study)

1.1.1 11a. Permutation

QUESTION: How does the difference in conversion rates for the permuted data compare to the conversion rate difference we saw in the original observed sample? What might that imply regarding our hypotheses?

ANSWER: Though both are small, the observed rate difference is about 8 times larger than the rate difference in the permuted data. This might lead us to think we should reject the null- it seems unlikely that the difference in rates for the control and experimental groups is due to random chance. However, we CANNOT make any conclusive statements based on a single permutation test statistic.

Note: your permutation conversion rate difference may not match the one shown here if you ran the `permutation` function multiple times, which is fine. In almost all cases, you should see a permutation conversion rate difference that is several times smaller than the observed rate difference.

1.1.2 11b. Simulation

1.1.3 11c. P-Values, T-Tests, and Standard Error

EXERCISE: Calculate the empirical p-value.

- Create a DataFrame `more_than_original` that contains all the rows of `differences` where `rate_difference` is greater than or equal to `observed_diff`. This code has been given.
- Get the count of items in the `more_than_original` DataFrame and the count of items in the `differences` DataFrame. Hint: the DataFrame attribute `shape` is helpful- remember, it's called using dot notation and doesn't have parentheses after since it's an attribute, not a function.
- Divide `more_than_count` by `total_diffs_count` to get the empirical p-value.

In [14]:
```
#  A DataFrame of values in the empirical distribution that are at least as large
more_than_original = differences[differences["rate difference"] >=observed_diff]

# the number of rows in more_than_original
more_than_count = more_than_original.shape[0]
# the number of rows in differences
total_diffs_count = differences.shape[0]

# the empirical p value
empirical_p = more_than_count / total_diffs_count
empirical_p
```

Out[14]: 0.0

QUESTION: A statistically significant p value is conventionally defined as less than or equal to 0.05, and a highly significant p value is conventionally defined as less than or equal to 0.01.

- Is the p-value we see significant?
- Given the p-value *alone*, should we recommend that Taskbella should continue the ad campaign? If not, what other information (perhaps from the first notebook) would we need, and what should our recommendation be?

ANSWER: Yes: the empirical p value is so small that our computer rounds it off to 0.0, meaning it is highly significant. No, we cannot make a recommendation on the p-value alone. The p-value tells us that the observed difference in conversion rates between the experimental and control groups is highly unlikely to happen by random chance. However, it is possible to have a highly significant difference between the two groups AND still not generate enough profits from the campaign to justify the costs. Furthermore, there are some other checks that we as data scientists would want to do before making a recommendation, including:

- make sure the test and control groups were properly randomized (i.e. that both groups are representative of the overall population of customers)
- check that the control group size was large enough to justify comparing it to the experimental group

In [2]:
```
import gsExport
gsExport.generateSubmission("09Machine-Learning-SOLUTIONS.nbconvert.ipynb")
```

```
[NbConvertApp] WARNING | Config option `template_path` not recognized by `NotebookExporter`
[NbConvertApp] Converting notebook 11AB-Test-SOLUTIONS.ipynb to notebook
[NbConvertApp] Writing 41520 bytes to 11AB-Test-SOLUTIONS.nbconvert.ipynb
```