0.0.1 Question 1a

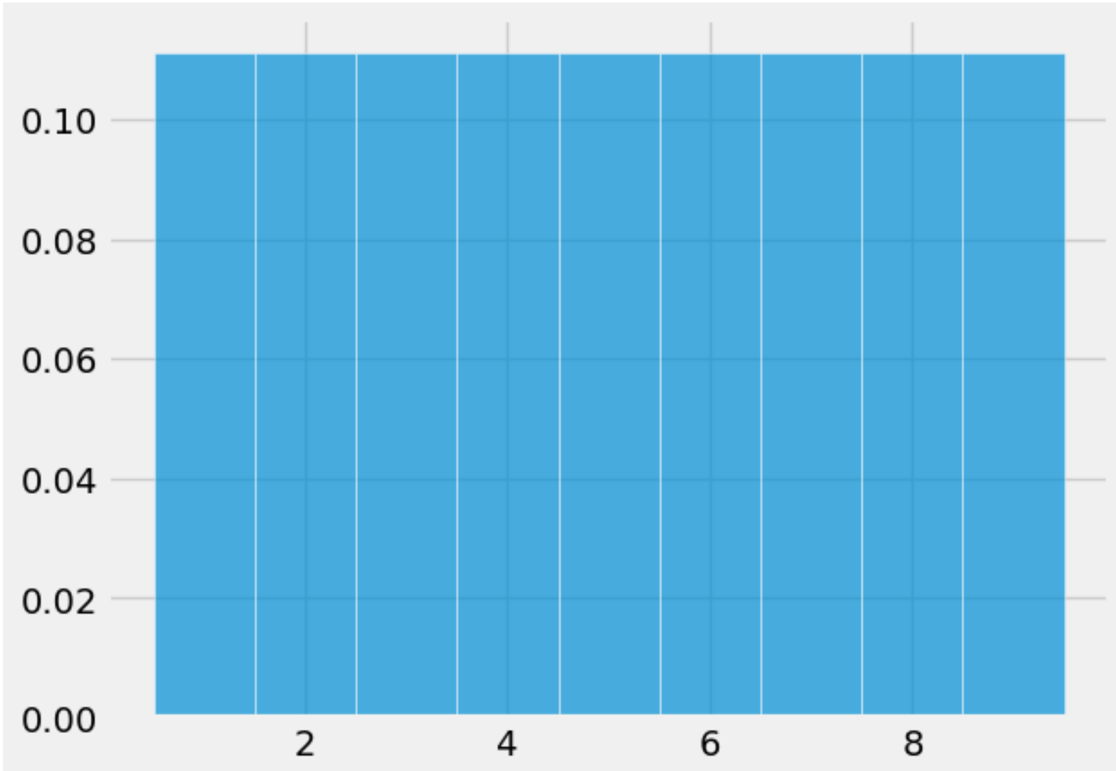
Define a function `integer_distribution` that takes an array of integers and draws the histogram of the distribution using unit bins centered at the integers and white edges for the bars. The histogram should be drawn to the density scale, and opacity should be 70%. The left-most bar should be centered at the smallest integer in the array, and the right-most bar at the largest.

Your function does not have to check that the input is an array consisting only of integers. The display does not need to include the printed proportions and bins. No title or labels are required for this question.

If you have trouble defining the function, go back and carefully read all the lines of code that resulted in the probability histogram of the number of spots on one roll of a die. Pay special attention to the bins.

Documentation: `plt.hist()` link

```
In [57]: def integer_distribution(arr_ints):
             unit_bins = np.arange(min(arr_ints)-0.5, max(arr_ints)+1.5, 1)
             plt.hist(arr_ints, bins=unit_bins, ec='white', density=True, alpha=0.7)
         faces = range(1, 10)
         integer_distribution(faces)
```


0.10
0.08
0.06
0.04
0.02
0.00
2
4
6
8

0.0.2 Question 1c

Before we write any code, let's review the idea of hypothesis testing with the permutation test. We first simulate the experiment many times (say, 10,000 times) through random permutation (i.e. without replacement). Assuming that the null hypothesis holds, this process will produce an empirical distribution of a predetermined test statistic. Then, we use this empirical distribution to compute an empirical p-value, which is then compared against a particular cutoff threshold in order to accept or reject our null hypothesis.

In the below cell, answer the following questions: * What does an empirical p-value from a permutation test mean in this particular context of birthweights and maternal smoking habits? * Suppose the resulting empirical p-value $p \leq 0.01$, where 0.01 is our p-value cutoff threshold. Do we accept or reject the null hypothesis? Why?

1. The empirical p-value means out of the many permutation tests, how many times that we see birthweights difference between random samples is larger (or smaller in the negative case) than the observed mean difference between moms with or without maternal smoking habits.
2. With empirical p-value lower than our p-value cutoff threshold, we reject the null hypothesis and accept the alternative hypothesis. It means the chance of random sampling matching the observed difference is lower than 1%. Hence the result is statistically significant.

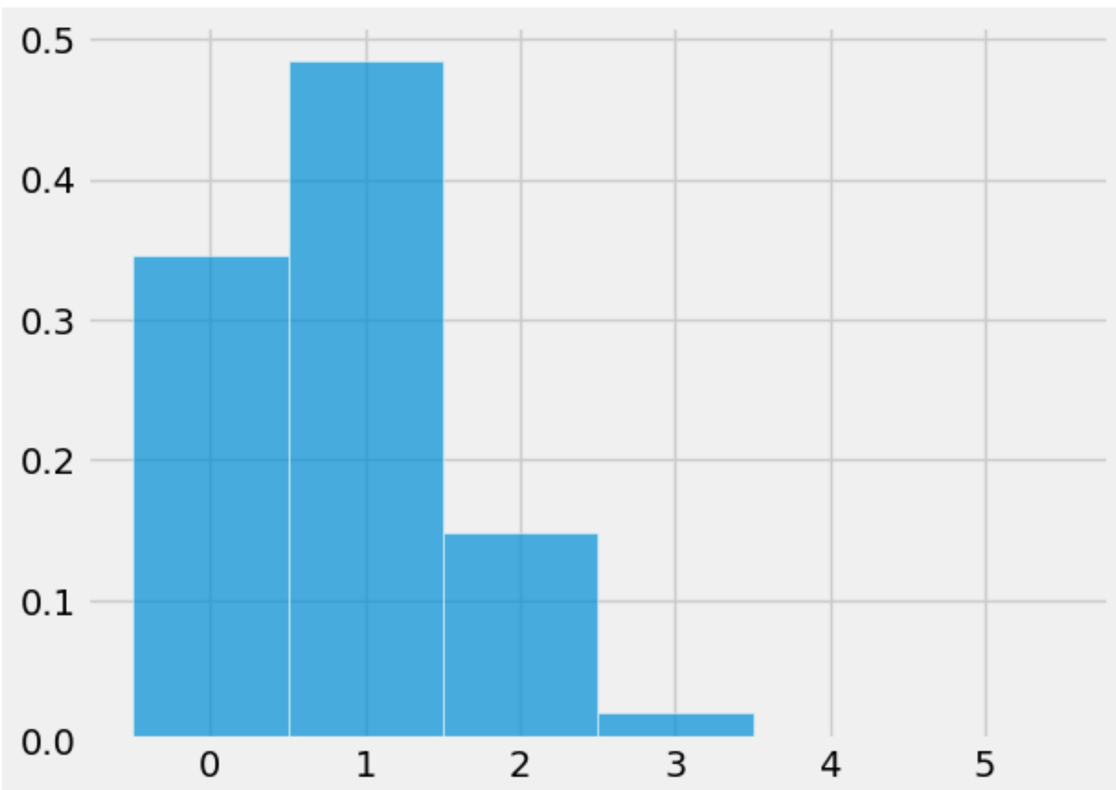
0.0.3 Question 1e

The array `differences` is an empirical distribution of the test statistic simulated under the null hypothesis. This is a prediction about the test statistic, based on the null hypothesis.

Use the `integer_distribution` function you defined in an earlier part to plot a histogram of this empirical distribution. Because you are using this function, your histogram should have unit bins, with bars centered at integers. No title or labels are required for this question.

Hint: This part should be very straightforward.

```
In [60]: integer_distribution(differences)
```



0.0.4 Question 1g

Based on your computed empirical p-value, do we accept or reject the null hypothesis? Be sure to include a reasonable p-value cutoff threshold, if any.

I reject the null hypothesis with a p-value of 0.01. The result is statistically significant.