NYC Squirrel Behavior Analysis

Abstract

The purpose of this paper is to investigate whether and in what ways urbanization has impacted Squirrel foraging behavior, using a dataset that focuses on Eastern Gray Squirrels (*Sciurus carolinensis*) in Central Park, New York City. We used data from the 2018 Central Park Squirrel Census to explore foraging observations in the context of broader national data as identified by Parker et. al (2014) and Thompson (1976). We then did more specific analysis to look for lurking variables and impacts of squirrel foraging as it pertains to the internal geographies and ecology of parks. We hypothesized that urban trends for squirrel foraging patterns - such as foraging frequency and hours of activity - would be replicated in this Central Park sample and that intra-park habitat would impact foraging behavior. Our conclusions showed that there was a difference in foraging patterns between Eastern Gray Squirrels in Central Park vs. other urban areas; however, we did not find any statistically significant results correlating intra-park geography and squirrel foraging patterns.

Introduction and Background

In our fast-paced modern world, we rarely sit down and truly think about how rapid modernization and globalization has impacted nature. In New York's central park there exists a population of Eastern Gray Squirrels (*Sciurus carolinensis*), who have gone through behavioral changes that helped them adapt to city living. In this paper, we want to find out the how and the why. These questions are of importance as they demonstrate the extent to which we have affected the natural order, causing a host of issues for existing populations and food chains.

Eastern gray squirrels are a naturally diurnal species, which means they have a circadian rhythm similar to ours, awake during the day and asleep at night. Studies have found that while wild squirrels display foraging hours in the morning to the late afternoon during the fall season (our season of interest), urban squirrels tend to remain active throughout the day (Thomspon 1976, Parker 2005). City living has major effects on behavior, and the New York metropolitan area has high rates of light pollution. Songbirds in major cities are said to sleep less and have a disrupted circadian rhythm compared to songbirds that lived in forests or rural areas (Dominoni et al., 2013). Circadian biology impacts every mammalian process down to a cellular level, and can have major impacts on feeding cycles (Pickel and Sung 2020). For example, it is suggested that severe modifications to circadian rhythm can create non-rhythmic eating patterns that result in equal feeding in light and dark periods of a 24 hour cycle (Pickel and Sung 2020).

Finally, it is important to understand the context and relevance around engagement between humans and squirrels in urban areas. To start, the ecological care and upkeep for Central Park is extensive and well funded: Central Park Conservancy has a 74 million dollar annual budget invested in care and upkeep for the park. (The Central Park Conservancy, n.d.) The non-profit works in conjunction with NYC Parks. Secondly, deep engagement with the park has meant that not only is the park itself cared for, but there is also significant energy put forth to ensure that people are engaging with the park and its wildlife in sensitive ways. One example of this is WildlifeNYC, launched in 2016 to raise awareness about wildlife in NYC, this campaign discourages the feeding of wild animals, including Eastern Gray Squirrels, in New York parks (WildlifeNYC, 2021). While there is no data on the efficacy of this campaign, studies have shown that when presented within the right framework, public awareness campaigns can change human behavior around targeted topics (Seymour 2018).

Methods

Data and Variables

This dataset was organized by a research team with data collected by volunteers. It contains observational data on 41 variables of the gray squirrels in NYC's central park. We combined our data set with two additional geographic data sets: the first one overlayed park permit zoning area to each squirrel observation location ("AreaType"), and the second a qualitative binary conservation data that marked whether the squirrel was observed in a protected area or not. Beyond that, we focused on the variables "shift" (AM/PM), "foraging" (True/False), and "status" (nature preserves and NA - all other parts of the park) to examine the squirrels foraging behaviors, and to explore whether foraging behaviors changed in Natural Preserves or Lawns as compared to the overall park.

Exploratory Data Analysis

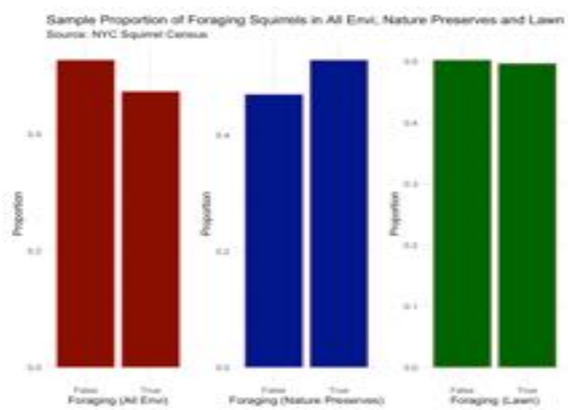


Figure 1: Squirrels were seen foraging for approximately 43% of the time in Central Park, 53% of the time in nature preserves, and 49% of the time on the lawn.

Figure 2: In the entire park, more squirrels are observed during the PM shift than the AM shift. Roughly 50% of the squirrels were foraging in the PM, and inversely, about 50% of squirrels were seen foraging in the AM. Additionally, we see that squirrels in Nature Preserves were seen foraging more often than those in the park at large (not delineated by geographic area), and the gap between the proportion of squirrels foraging and of squirrels not foraging in the AM is slightly more larger than that of the PM times. In the lawns, during the PM shift, approximately the same amount of squirrels were seen foraging and not foraging; however, in the morning, slightly fewer squirrels were seen foraging.

Analytic Methods

Our data lies at the intersection of a census and a sample, so we decided to approach this data set in two different ways: as a sample of urban squirrels in North America, and as its own population. To start, we interpreted our data as an urban sample of squirrels and then compared the proportion of squirrels foraging from our data to that of Parker et. al (2014) in Baltimore - standing in for the population of U.S. urban squirrel foraging proportion - using a two-sample z-test. Then, diving deeper into the squirrels' foraging behavior: the urban squirrels from Baltimore, MD and Toronto, Ontario actively foraged throughout the day, our sample showed no preference between the mornings and the afternoons, we decided to do a two sample z-test to see if this lack of preference is also reflected in our Central Park squirrel sample (Parker et. al, 2014; Thompson, 1997). Since, we overlaid detailed geographic information for each entry onto the data, two-sample z-tests were performed two more times, first within nature preserves and then within lawns, to see if foraging behavior throughout the day in these two geographic subsets were different.

When treating the census of squirrels in Central Park as a population, we were able to find out if any geographic subsets (repetitions = 500 and 5000) had statistically different foraging proportions to the larger census population. This was done to pretend that we collected the data a lot of times instead of just one. With only one sample, we were putting more emphasis on the behavior of specific squirrels within that trial instead of having many trials, which would de-emphasize the specific observations and highlight trends instead. By sampling multiple times, we are removing the assumption that that 1 trial of 348 squirrels represents overall squirrel preference for that location in terms of foraging, we will be looking at the mean p-hat of the 5000 trials with different combinations of 35 observations taken from the initial sample, which will be much closer to the true p-hat for squirrel foraging in that location. Two one-sample z-tests were used to compare the mean proportion of squirrels foraging in these subset samples to the proportion of squirrels foraging in the population, which showed us if squirrels really preferred foraging at a specific location over another. Two more one-sample z-tests were used to compare the mean proportion of squirrels foraging in these subset samples to the proportion of squirrels foraging on these subsets in the original census.

Results

Null Hypothesis	p-value	z-score	Result
$\hat{p_{NYC}} = \hat{p_{MD}}$	0	20.35	H_0 rejected
$\hat{p_{AM}} = \hat{p_{PM}}$	0.001871	-3.11	H_0 rejected
$\hat{p_{nature\ preserve}} = \mu_{park}$	0.7	0.477	Fail to reject
$\hat{p_{lawn}} = \mu_{park}$	0.2846	1.07	Fail to reject
$\hat{p_{AM\ nature\ preserve}} = p_{PM}$	.14986	-1.44335	Fail to reject
$\hat{p_{AM\ lawn}} = \hat{p_{PM\ lawn}}$	.0198	2.33333	Fail to reject

When we treated our Central Park squirrels as a sample, they were observed foraging ($\hat{p}$ = 0.473) almost 50% more often than the population data collected by Parker et. al. from 6 urban parks in Baltimore, MD suggested ($\hat{p}$ = 0.252). Our results provide evidence (z = 20.4, p = 0) at the 1% significance level that squirrels forage in Central Park differently than squirrels in the Maryland sample. That said, we did not find significant differences in foraging when looking specifically at the intra-park ecological habitats compared to overall park foraging (preserve: $z = 0.477$, $p = 0.7$; lawn: $z = 1.07$, $p = 0.2846$). When comparing the lawn and preserve samples to their sampling distributions, we failed to reject the null hypothesis and were able to infer that the samples were unbiased. Central Park squirrels also seem to forage more actively in the afternoon (PM sighting session) than the morning ($z = -3.11$, $p = 0.0019$), unlike the two samples from Toronto and Baltimore that showed no skew. That said, when we calculated the difference in foraging between AM and PM in specific intra-park terrains at a 1% significance interval, we did not find significant differences (Preserve: $z = .1.44335$, $p = .14986$; Lawn: $z = 2.33333.$, $p = .0198$).

Discussion

There were two aspects to our study that depended on the unique traits of our data set. In the first part of our analysis, we treated our data set as though it was a sample from the larger population of Eastern Gray Squirrels (*Sciurus carolinensis*) in North America. In the second part of our analysis, we were able to treat the data set as a population due to it being a "census" and extract samples. We then compare the proportion our sampling distribution proposed with that of the population to evaluate the results and how biased they are.

Firstly, we had two significant results from our data testing, both when comparing our data to other studies. When comparing time spent foraging, we found that the NYC Central Park squirrels spent significantly more time foraging than the Maryland sample. One explanation for this could be decreased levels of food or ideal food sources, such as nuts and seeds, which would force the squirrels to spend more time foraging than they otherwise would. (Petrucci, 1993) The percentage of time spent foraging by Central Park Squirrels, while significantly higher than that of squirrels in Maryland parks, seems to be in line with foraging of rural squirrels or squirrels in areas with high tree density (forests). (Shuttleworth, 2000) The marked difference in time spent foraging could be due to human interference, such as feeding of squirrels and human food being left behind after visitation. This seems even more plausible considering active involvement of NYC Parks and The Central Park Conservancy and the WildlifeNYC campaign started in 2016. (Central Park Conservancy, 2017) Additionally, we found that there was significantly more foraging activity in the evening/afternoon vs the morning. This could potentially be due to the high levels of light pollution in New York City which in turn affect circadian rhythms. The census data suggests increased squirrel feeding in the afternoon/early evening, which could mean a disturbance in the squirrel's circadian rhythm, pushing back feeding times to non-light hours.

One thing that is important to note in our results is the importance of sample sizes on significance in hypothesis tests. The higher a sample size is, the easier it is to have statistical significance; however, this does not always correlate with significant differences. Because our census sample is large ($n = 3042$), it is easier to have significant results, which we have observed; however, in the tests that included samples with smaller sample sizes, such as the lawn and nature preserve tests, we saw no statistical significance. Additionally, the insignificance of results in inter-park statistical tests could also be due to non-significant variation of foraging behavior within Central Park. This analysis would need to be repeated in other parks for further conclusions to be drawn about how inter-park habitat differences impact squirrel foraging behavior.

Our analysis has other limitations as well. Firstly, as our dataset is collected by a team of volunteers, it is prone to errors that would reduce the reliability of our downstream analysis. Also, the assumptions we made, such as assuming that the NYC Census sample was a population in itself, could be false, invalidating our conclusions done through cluster sampling.

Our results highlight the variation between our population and the Maryland population, while finding no differences between different areas inside Central Park, we hope to motivate further research on how certain elements of urban habitats (light pollution, for example) can affect squirrels behavior. We are also interested in seeing how our squirrel analysis could be replicated for another species. Hopefully, this line of research would continue to be of interest and provide useful insights into urbanization and ecology.

References

About WildlifeNYC (2021). Retrieved from
https://www1.nyc.gov/site/wildlifenyc/about/about-wildlifenyc.page

Dominoni, D. M., Helm, B., Lehmann, M., Dowse, H. B., & Partecke, J. (2013, July 22). Clocks for the city: circadian differences between forest and city songbirds. *Proceedings of the Royal Society B: Biological Sciences, 280*(1763).

Parker, T. S., Gonzales, S. K., & Nilon, C. H. (2014). Seasonal comparisons of daily activity budgets of gray squirrels (sciurus carolinensis) in urban areas. *Urban Ecosystems, 17*(4), 969-978.

Pickel, L., & Sung, H. K. (2020). Feeding rhythms and the circadian regulation of metabolism. *Frontiers in nutrition*, *7*.

Thompson, D. C. (1977, July). Diurnal and seasonal activity of the grey squirrel (Sciurus carolinensis). *Canadian Journal of Zoology, 55*(7), 1185-1189.

Walker II, W. H., Walton, J. C., Courtney Devries, W. A., & Nelson, R. J. (2020, January 23). Circadian rhythm disruption and mental health. *Nature: Translational Psychiatry, 10*(28).

Seymour J. (2018). The Impact of Public Health Awareness Campaigns on the Awareness and Quality of Palliative Care. Journal of palliative medicine, 21(S1), S30–S36.

Petrucci, S. T. (1993). Observations of the Foraging Behavior of Squirrel Monkeys (Saimiri sciureus boliviensis) at the Tulsa Zoological Park. *Bios*, 64(3), 64-71.

Shuttleworth, C. M. (2000, September). The foraging behaviour and diet of red squirrels Sciurus vulgaris receiving supplemental feeding. *Wildlife Biology* 6(4), 149-156.

The Central Park Conservancy (2017, January). Winter Stories from Central Park's Wildlife. Retrieved from https://www.centralparknyc.org/articles/winter-stories-wildlife

The Central Park Conservancy. Governance. https://www.centralparknyc.org/governance

Appendix

Links to Dataset:

The New York Central Park squirrel census can be found at https://catalog.data.gov/dataset/2018-central-park-squirrel-census-squirrel-data, the Park Permit Data at https://data.cityofnewyork.us/Recreation/Parks-Permit-Areas-Map/fc69-ufed, and the Conservation Data Set at https://data.cityofnewyork.us/Environment/Forever-Wild-Preserves-and-Natural-Areas/u32x-nkau.

Hypothesis Test 1: Squirrels' time spent foraging

Hypotheses:

$H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$. There are no differences between the proportion of time the squirrels spent foraging in the central park sample and Dr. Parker's sample.

$H_A = \hat{p_1} - \hat{p_2} \neq 0$. There is a difference between the proportion of time the squirrels spent foraging in the central park sample and Dr. Parker's sample.

Checking Assumptions:

- Independent assumption: It is reasonable to assume that each squirrel's behavior doesn't depend on another squirrel's behavior.
- Independent group assumption: Our central park squirrels is independent from Dr. Parker's squirrels from 6 parks in Baltimore, MD.
- 10% Condition: According to *Annual City Parks Data Released by The Trust for Public Land*, there are 22,493 city parks in the United States. Since we are only looking at two parks, we can safely assume that we cover less than 10% of the urban squirrel population.
- Success/Failure Condition: Our central park sample has 1603 failures and 1439 successes, while Dr. Parker sample has approximately 377 failures and 127 successes.

Therefore, we used 2 sample hypothesis testing (z-test) to find out of there is a difference in Dr. Parker squirrels and our squirrels foraging.

Calculating the test statistic:

- $\hat{p_1}$ = our sample's foraging proportion (0.473) with a sample size of n_1 = 3042
- $\hat{p_2}$ = Dr. Parker's sample's proportion (Fall 2004, 25.2 %), n_2 = 504.

SE = $\sqrt{\frac{0.473*(1-0.473)}{3042} + \frac{0.252*(1-0.252)}{504}} = 0.02135$

z-score = $\frac{0.473-0.252}{0.02135} = 10.35$

$P(|z| > 20.35) = 2 * (1 - pnorm(10.35)) = 0$

We reject the Null Hypothesis: there is a statistically significant difference between the proportion of Dr. Parker's squirrel sample foraging and that of our sample's foraging.

Hypothesis Test 2: Squirrels' foraging behavior throughout the day

Hypotheses:

$H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$: Just like Dr. Thompson's sample, we have the same proportion of squirrels foraging in the AM and in the PM.

$H_A = \hat{p_1} - \hat{p_2} \neq 0$: We don't have the same proportion of squirrels foraging in the AM and in the PM.

Checking Assumptions:

- Independent assumption: It is reasonable to assume that each squirrel's behavior doesn't depend on another squirrel's behavior within each data set.
- Independent group assumption: Because AM and PM are two different times of the day, and each data point within those time frames is a different squirrel, the two groups are independent.
- 10% Condition: We can be confident that our sample has fewer than 10% of the North American squirrels population.
- Success/Failure Condition: In the AM sample, we have 751 non-foraging squirrels and 602 foraging squirrels, and in the PM sample, we have 852 non-foraging squirrels and 837 foraging squirrels.

Therefore, we used 2 sample hypothesis testing (z-test) to find out if there is a difference between AM and PM foraging behavior in the fall.

Calculating the test statistic:

- $\hat{p_1} = 0.44,\ n_1 = 1353$ for the AM Shift
- $\hat{p_2}$ 0.496, $n_2 = 1689$ for the PM Shift

SE = $\sqrt{\frac{0.44*(1-0.44)}{1353} + \frac{0.496*(1-0.496)}{1689}} = 0.018$

z-score = $\frac{0.44-0.496}{0.018} = -3.11$

- $P(|z| > -3.11) = 2 * pnorm(-3.11) = 0.001871$

H_0 is rejected, there is a difference between AM and PM foraging behavior, which means the squirrels' foraging times are affected by city living. The bar chart made shows accurate proportions.

Hypothesis Test 3: Squirrels' foraging behavior throughout the day on lawns

Hypotheses:

$H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$: Just like Dr. Thompson's sample, we have the same proportion of squirrels foraging on the lawn in the AM and in the PM.

$H_A = \hat{p_1} - \hat{p_2} \neq 0$: We don't have the same proportion of squirrels foraging on the lawn in the AM and in the PM.

Checking Assumptions:

- Independent assumption: It is reasonable to assume that each squirrel's behavior doesn't depend on another squirrel's behavior within each data set.
- Independent group assumption: Because AM and PM are two different times of the day, and each data point within those time frames is a different squirrel, the two groups are independent.
- 10% Condition: We can be confident that our sample has fewer than 10% of the North American squirrels population.
- Success/Failure Condition: In the AM sample, we have 75 non-foraging squirrels and 70 foraging squirrels, and in the PM sample, we have 100 non-foraging squirrels and 103 foraging squirrels.

Therefore, we used 2 sample hypothesis testing (z-test) to find out if there is a difference between AM and PM foraging behavior in the fall.

Calculating the test statistic:

$\hat{p_1}$ = AM Shift = 0.4827586 $\hat{p_2}$ = PM Shift = 0.5073892

- $H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$
- $H_A = \hat{p_1} - \hat{p_2} \neq 0$
- $P(|z| < -0.4531) = 2 * pnorm(-0.4531) = .65272$

Because p = .65272, we fail to reject the null hypothesis at p < .01. This means that there is not a significant difference between squirrel foraging in the AM vs. PM on Lawns.

Hypothesis Test 4: Squirrels' foraging behavior throughout the day in nature preserves

Hypotheses:

$H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$: Just like Dr. Thompson's sample, we have the same proportion of squirrels foraging in the pature preserve in the AM and in the PM.

$H_A = \hat{p_1} - \hat{p_2} \neq 0$: We don't have the same proportion of squirrels foraging in the nature preserve in the AM and in the PM.

Checking Assumptions:

- Independent assumption: It is reasonable to assume that each squirrel's behavior doesn't depend on another squirrel's behavior within each data set.
- Independent group assumption: Because AM and PM are two different times of the day, and each data point within those time frames is a different squirrel, the two groups are independent.
- 10% Condition: We can be confident that our sample has fewer than 10% of the North American squirrels population.
- Success/Failure Condition: In the AM sample, we have 75 non-foraging squirrels and 70 foraging squirrels, and in the PM sample, we have 100 non-foraging squirrels and 103 foraging squirrels.

Therefore, we used 2 sample hypothesis testing (z-test) to find out if there is a difference between AM and PM foraging behavior in the fall.

Calculating the test statistic:

- Used 2 sample hypothesis testing (z-test) to find out if there is a difference between AM and PM foraging behavior in the fall in nature preserves. The paper states that AM and PM foraging behavior should be the same, 5 hours in AM and 5 hours in PM times.

$\hat{p_1}$ = AM Shift = 0.5404255

$\hat{p_2}$ = PM Shift = 0.5180180

- $H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$
- $H_A = \hat{p_1} - \hat{p_2} \neq 0$
- $P(|z| < 0.4797) = 2 * (1 - pnorm(0.4797)) = .63122$

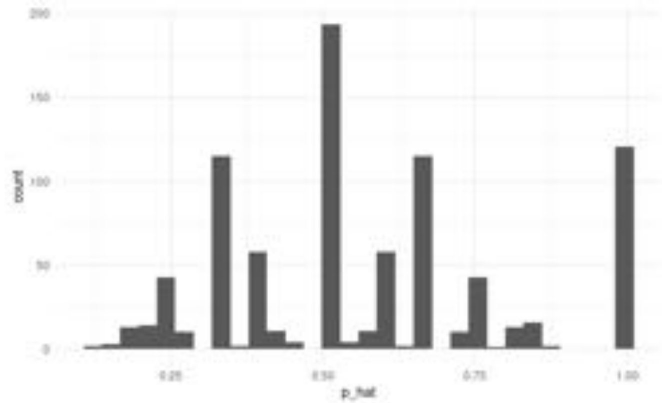
Because p = .63122, we fail to reject the null hypothesis at p < .01. This means that there is not a significant difference between squirrel foraging in the AM vs. PM in nature preserves.

Hypothesis Test 5: Comparing $\hat{p}$ of foraging behavior in lawns to $\hat{p}$ of foraging behavior in overall sample

Hypotheses:

- $H_0 : \hat{p_1} = \mu_2$: There is no difference between the mean proportion of the 5000 lawn samples we've drawn and the Census foraging proportion.
- $H_A : \hat{p_1} \neq \mu_2$: There is a difference between the mean proportion of the 5000 lawn samples we've drawn and the Census foraging proportion.

Checking Assumptions:



- Independence assumption: It is reasonable to assume that each squirrel's behavior doesn't depend on another squirrel's behavior within each data set.
- Independent group assumption: While the Lawn sample is extracted from the squirrel's census, the Lawn samples drawn are approximately 1% of the squirrel census, so we can assume that those two are independent samples
- 10% condition: There are a total of 348 squirrels in lawns in Central Park, and we are drawing samples of size 10.
- Success/Failure Condition: In the median sample, we have approximately 50% of squirrels foraging, which means that 15 foraging squirrels and 15 non-foraging squirrels. In the population, we have 1603 non-foraging squirrels and 1439 foraging squirrels.

Calculating the test statistic:

Hypothesis test comparing p_hat of foraging behavior on lawn of cluster sample to p_hat of foraging behavior on lawn of overall sample

$\hat{p_1}$ = Lawn foraging proportion (0.5) $n_1 = 348$

$\hat{p_2}$ = Cluster Lawn foraging proportion (0.57) *acting as P* $n_2 = 35$

$H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$ $H_1 = \hat{p_1} - \hat{p_2} \neq 0$

P(|z| < 2.33333.)

P-value = .0198

At p < .01, we fail to reject the null hypothesis. The overall sample proportion of foraging on Lawns not statistically different from the cluster sample proportion of foraging on lawns. This suggests that our nature preserve sample is random.

Hypothesis Test 6: Comparing $\hat{p}$ of foraging behavior in nature preserves to $\hat{p}$ of foraging behavior in overall sample

Hypotheses:

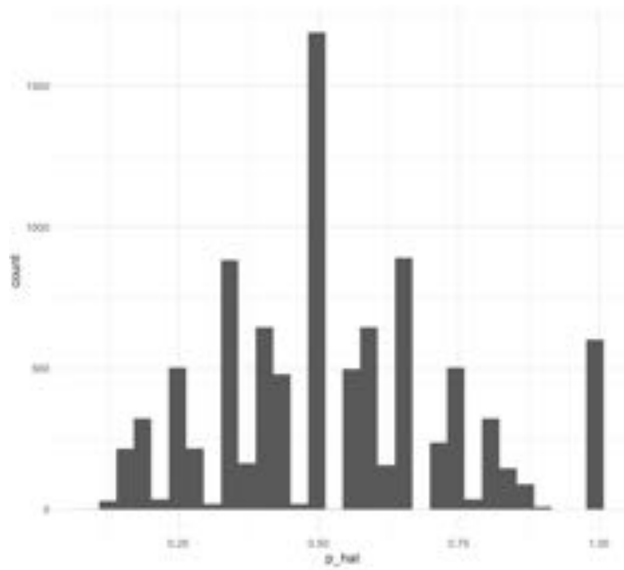
$H_0 = \hat{p_1} = \mu_2$

$H_A = \hat{p_1} \neq \mu_2$

Checking Assumptions:

- Independence assumption: It is reasonable to assume that each squirrel's behavior doesn't depend on another squirrel's behavior within each data set.

- 10% condition: There are a total of 457 squirrels in Natural Preserves in Central Park, and we are drawing samples of size 35.
- Success/Failure Condition: In the median sample, we have approximately 50% of squirrels foraging, which means that 15 foraging squirrels and 15 non-foraging squirrels. In the population, we have 1603 non-foraging squirrels and 1439 foraging squirrels.



Calculating the test statistic:

$\hat{p_1}$ = Nature Preserve foraging proportion (0.53)

n_1 = 457 $\hat{p_2}$ = Cluster Nature Preserve foraging proportion (0.567) *acting as P*

$H_0 = \hat{p_1} - \hat{p_2} = 0 = d_0$

$H_1 = \hat{p_1} - \hat{p_2} \neq 0$

P(|z| <-1.44335)

P-value = .14986

At p < .01, we fail to reject the null hypothesis. The overall sample proportion of foraging in nature preserves is not statistically different from the cluster sample proportion of foraging in nature preserves. This suggests that our nature preserve sample is random.