1. Download the class data and import the file into R with Rcmdr. Use Rcmdr to estimate the following descriptive statistics (1 pt). Note that all these descriptive or summary statistics are not necessarily produced in Rcmdr, so you will need to problem solve to determine the answers a through e. Print or reproduce your Rcmdr results and highlight a-e for our reference.

Q1 1

> numSummary(Habitat_sel[,"Prop_Forest_Interior", drop=FALSE], statistics=c("mean",

+   "sd", "se(mean)", "IQR", "quantiles", "cv", "skewness", "kurtosis"),

+   quantiles=c(0,.25,.5,.75,1), type="2")

  mean   sd   se(mean)   IQR   cv skewness kurtosis   0%   25%   50%

55.5556 3.194661 0.3194661 3.9375 0.05750385 0.1349101 0.5345027 47.67 53.63 55.535

  75% 100% n

57.5675 64.67 100

a. Mean = 55.5556

b. Median = 55.535

c. Variance = (standard deviation)^2 = 10.20

d. Standard deviation = 3.194661

e. Standard error (se) = 0.3194661

2. Use R Cmdr to produce a histogram displaying the distribution (make sure the axes are labelled and that you have a proper figure caption.

Q2 **0.75**

'Y' axis label is "Frequency" 'X' axis label is not "expectation" but the "proportion" of time a species of bird spent in forest interior relative to total available habitat
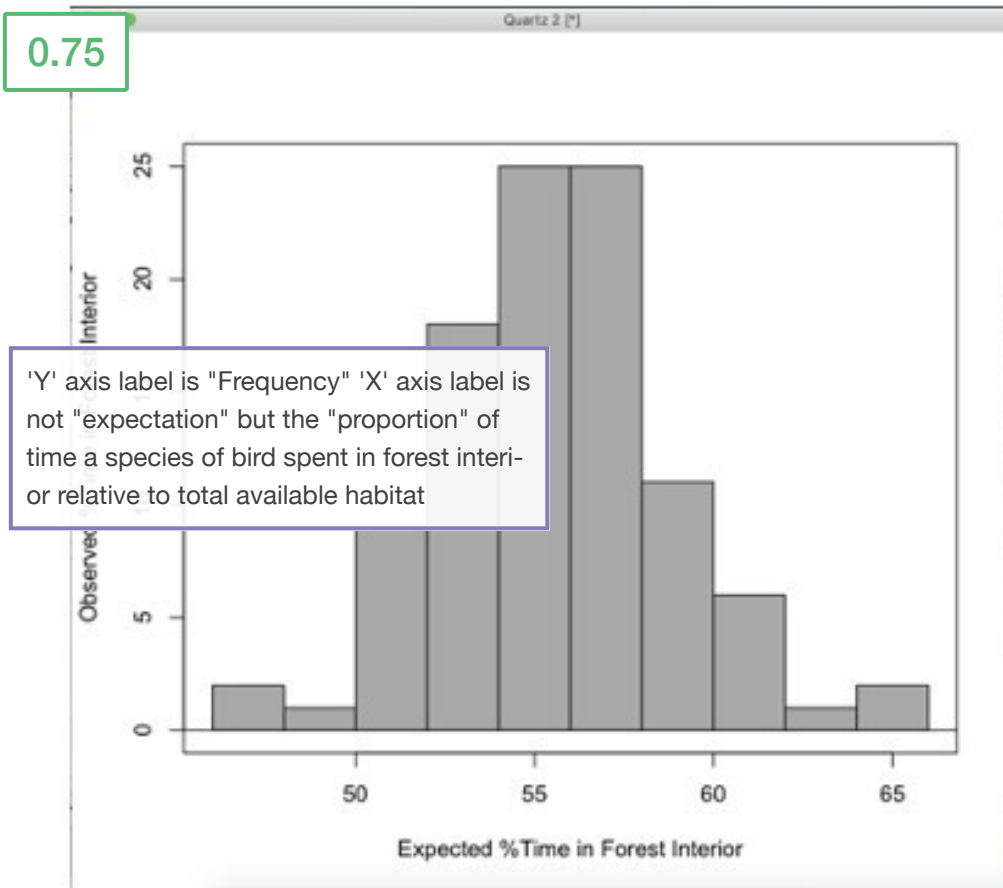
Figure 1. The Dataset Shows a Normal Distribution. This histogram shows that the observed distribution exhibits a normal distribution property. The distribution is centered, i.e., it's neither skewed to the right nor skewed to the left.

3. Use R Cmdr to test the assumption that your data are normally distributed. Reproduce results of the Shapiro Wilk's test (0.5 pt), K-S Lillefors test (0.5 pt), and a Q-Q plot (0.5 pt), and provide your R Cmdr output and plot. Again, make sure you have a figure caption and appropriate axis titles.

Q3 **1**

Shapiro-Wilk normality test	Lilliefors (Kolmogorov-Smirnov) normality test
data: Prop_Forest_Interior	data: Prop_Forest_Interior
W = 0.98973, p-value = 0.6426	D = 0.056665, p-value = 0.5977

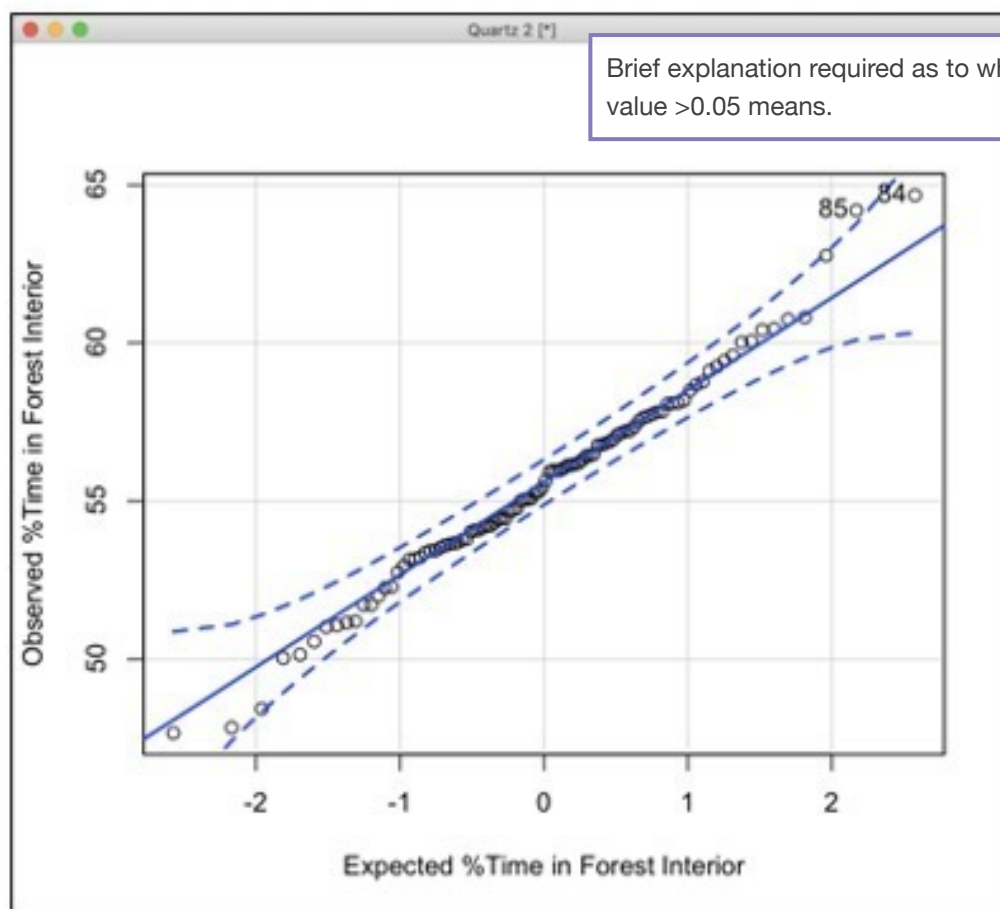Brief explanation required as to what a p value >0.05 means.

Figure 2. The Q-Q plot of the habitat selection data set shows a normal distribution plot. All the data points fall in the compatibility interval.

4. Based on your results in Q2 & Q3, are the sampled data normally distributed? Explain what
   you found based on these results. Does your data need to be transformed to meet the
   assumption of normality? If so, which transformation would you suggest? Provide an
   explanation in sentences. 1 pt for interpretation of Q-Q plot; 1 pt for what you found across
   Q2 & 3, and 0.5 pts for statement about transformation/assumption.

Q4   2.5

   - Base on the Q2 and Q3 results, the sampled data is normally distributed. The histogram
     in Figure 1 shows that the sampled data graph is bell-shaped and is almost perfectly
     symmetric about the middle (mean). The p-values of Shapiro-Wilks test and Lilliefors test
     are both greater than 0.05, which suggests that there is no difference between the
     sampled data distribution and the theoretically normal distribution. In addition, The Q-Q
     plot on Figure 2 shows the normality of the sampled data. In the Q-Q plot, the data
     points lie within the compatibility interval and although not perfectly straight, they show
     linearity because they follow along the straight blue curve. Thus, no transformation is
     needed as the sampled data already meets the criterion for normality.

5. What is your null hypothesis? (0.5 pt) What is your alternate hypothesis? (0.5 pt) What type of t-test would you use? Perform the t-test and report: what you did (1 pt); what you found (1 pt); and interpret your results ecologically (1 pt)

   a. The null hypothesis is $\mu_0 = 50$. This means that birds exhibit no selection for either forest interior or edge habitat.

Q5   **2.5**   The alternate hypothesis is $\mu_0 \ne 50$ (not equal to 50). This means that birds exhibit selection for either forest interior or edge habitat.

   c. The type of t-test that should be used is the single-sample t-test. To perform the single-sample t-test, I used R single-sample t-test function. I did this by simply going to Statistics > Means > Single-sample t-test. I set the null hypothesis as 50 and set the confidence level to 0.95. I also set the variable to "Prop_Forest_Interior. See Figure 3.
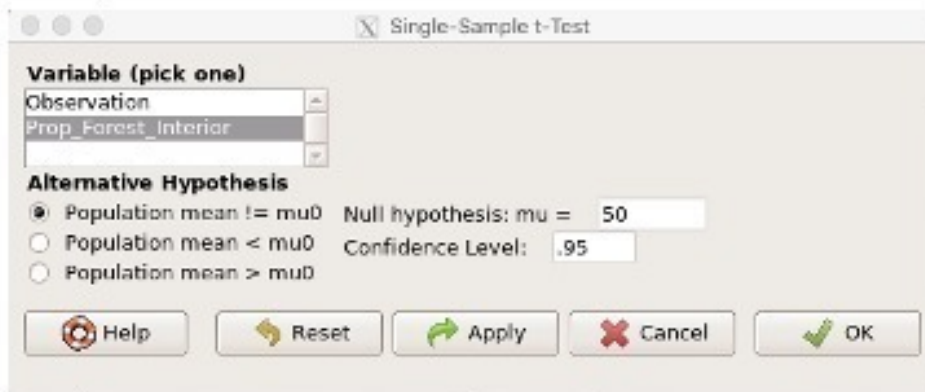

Figure 3. Single-sample t-test in R.

   d. The result from the single-sample t-test of the sampled data is shown below. The t-stat is equal to 17.39, the degrees of freedom (df) is 99, and the p-value of the sampled data is less than 2.2e-16. The sampled data's p-value is a lot less than the hypothetical p-value of 0.05.

```
> with(Habitat_sel, (t.test(Prop_Forest_Interior, alternative='two.sided', mu=50,
+ conf.level=.95)))
                        One Sample t-test
data: Prop_Forest_Interior
t = 17.39, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 54.92171 56.18949
sample estimates:
mean of x
 55.5556
```

e. Since the p-value of the sampled data is less than 0.05, the null hypothesis is rejected and the alternate hypothesis is accepted. It is safe to conclude that birds exhibit selection for either forest interior or edge habitat.
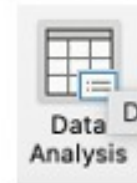
> Analysis unclear... Implies that the P-value indicating accepting the Null hypothesis which is incorrect. Review notes on P-values and hypothesis testing

6. Explain how you would go about drawing a random sample of 5 observations from this dataset (1 pt)

   - I used the "sampling" function in the Data Analysis button in Microsoft Excel. Then I entered "5" as the number of random samples. The obtained samples are placed in a column named "Observation Number". To obtain the corresponding Prop_Forest_Interior data, I used the VLOOKUP function of Excel.

Q6 | **1**

D	E	F	G	H
	OBSERVATION NUMBER	Prop_Forest_Interior_random		
RANDOM SAMPLE 1	94	58.1	MEAN	55.026
	79	53.64	STDEV	1.825960021
	41	55.05		
	95	53.67		
	57	54.67		
RANDOM SAMPLE 2	25	60.02	MEAN	56.056
	66	56.96	STDEV	3.167180134
	28	51.2		
	24	55.97		
	48	56.13		
RANDOM SAMPLE 3	23	51.07	MEAN	55.324
	100	60.45	STDEV	3.958905152
	34	58.5		
	33	53.44		
	56	53.16		
RANDOM SAMPLE 4	75	56.05	MEAN	57.514
	24	55.97	STDEV	1.53265456
	27	59.6		
	11	57.76		
	55	58.19		
RANDOM SAMPLE 5	12	55.05	MEAN	53.12
	35	51.16	STDEV	2.438759931
	64	50.04		
	76	55.65		
	46	53.7		

Q7 1

Highest mean (circled: 57.514 / 1.53265456)

Lowest mean (circled: 53.12 / 2.438759931)

D	E	F	G	H
RANDOM SAMPLE 6	21	51.72	MEAN	55.05
	52	57.84	STDEV	4.11921109
	29	50.56		
	7	54.72		
	53	60.41		
RANDOM SAMPLE 7	67	54.42	MEAN	54.056
	29	50.56	STDEV	2.737376481
	95	53.67		
	55	58.19		
	38	53.44		
RANDOM SAMPLE 8	39	54.22	MEAN	54.046
	44	52.76	STDEV	1.465377085
	9	56.49		
	90	53.53		
	14	53.23		
RANDOM SAMPLE 9	75	56.05	MEAN	55.932
	48	56.13	STDEV	1.006091447
	72	56.78		
	9	56.49		
	6	54.21		
RANDOM SAMPLE 10	9	56.49	MEAN	55.728
	32	53.8	STDEV	1.394478397
	99	54.76		
	43	56.43		
	69	57.16		

8. en you consider the mean +/- standard deviations for each of these two samples (i.e., the lowest mean and the highest mean), do they overlap? Why or why not? (1 pt)

Q8 | 1

- The standard deviations of these two samples do not overlap, they differ by approximately 1 unit. These 2 sample groups' standard deviations do not overlap because they are very distinct sample groups. In other words, they do not overlap because there is variability in the population that was selected.

9. What might you expect to have happened had your sample size increased from 5

Q9 2 vations per sample to 10 observations per sample? Why?

There will be a smaller difference between the standard deviations of the lowest mean and the highest mean because there is a greater precision and lesser uncertainty in a sample of 10 observations compared to 5 observations. The sample statistics will be a better representation of the population parameter for 10 observations as opposed to 5 observations because the larger the sample size is, the closer we are to the true value of the population mean.