GRADTDA 5402 SP25: PROJECT 2

FLIGHT RESPONSE OF THE PACIFIC BRANT

IZEMBEK LAGOON, ALASKA

Dheel, Newman, Siefker

INTRODUCTION

Increasingly given climate change and the impact that human development is having on the environment, policies are important for ensuring that we minimize the impact on protected habitats and wildlife. Threats to the environment are not limited to the most publicized factors such as pollution, deforestation, climate change and urbanization. Industrial machinery and transportation are drivers of noise pollution, another agent of stress for wildlife in exposed areas. To mitigate these risks in air transit, the Federal Aviation Administration has imposed a minimum altitude constraint of 2,000 feet when flying over marine sanctuaries. (https://sanctuaries.noaa.gov/flight/)

The following study is an analysis of data collected from a specific area, Izembek Lagoon, Alaska regarding a specific species of waterfowl, the Pacific brant. The data is the result of an observational study recording measurements of altitude for 464 helicopter flights. Lateral distance was calculated using maps of the flight path compared to the center of the flock. Researchers observed whether the documented flights elicited a response from the Pacific brant. While the type of response was not specifically identified in the data, responses include typical disruptions to normal patterns of the brant. As noted by NOAA, examples of these disruptions include but are not limited to increased stress and nest abandonment which could lead to issues with breeding. (https://sanctuaries.noaa.gov/flight/)

Key questions:

- What is the effect of helicopter altitude on the flight response of the Pacific brant?
- What is the effect of lateral distance on the flight response of the Pacific brant?

DATA EXPLORATION

The collected data contained 4 variables:

- *Observ:* *Numerical code identifying the flock*
- *Altitude:* *The altitude of the helicopter in units of 100m*
- *Lateral:* *The lateral distance of the helicopter to the center of each flock in units of 100m*
- *Flight:* *The categorical response variable indicating whether the observed flight generated a response. (1=yes, 0=no)*

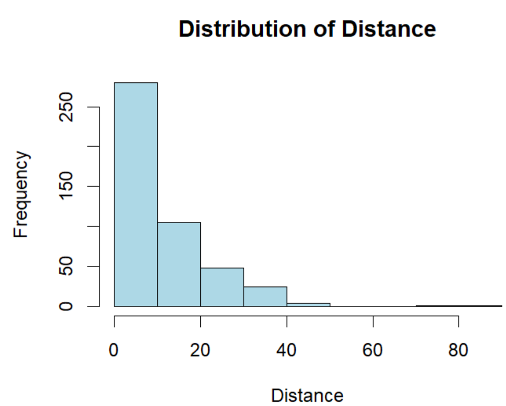
- Each attribute in the data is summarized in the table below:

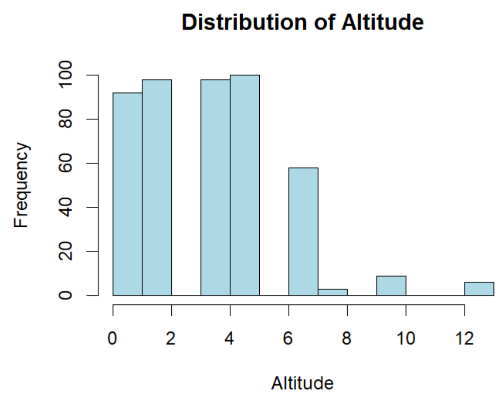
```
    observ          altitude
 Min.   :  1.0   Min.   : 0.910
 1st Qu.:116.8   1st Qu.: 1.520
 Median :232.5   Median : 3.050
 Mean   :232.5   Mean   : 3.284
 3rd Qu.:348.2   3rd Qu.: 4.570
 Max.   :464.0   Max.   :12.190
     lateral          flight
 Min.   : 0.00   Min.   :0.0000
 1st Qu.: 3.38   1st Qu.:0.0000
 Median : 6.60   Median :1.0000
 Mean   :10.97   Mean   :0.6142
 3rd Qu.:16.25   3rd Qu.:1.0000
 Max.   :80.47   Max.   :1.0000
```

For easier analysis we created two additional categorical variables for ease of analysis:

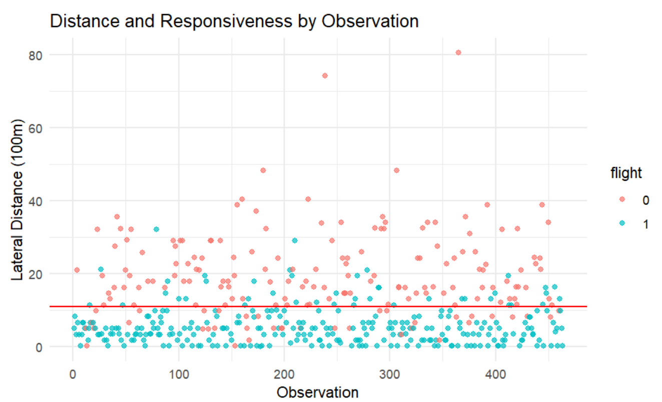
- *Lateral_cat:* *Grouping lateral distances into categories: <10,10-19, 20-29, >30*
- *Altitude_cat:* *Grouping altitude into categories: <3, 3-6, >6*

The following visualizations provided additional insight into the data.

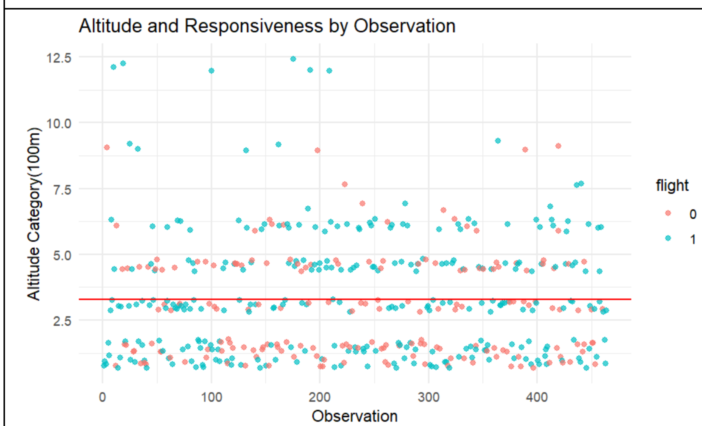
	The distribution of the lateral distances collected as part of the study were more frequently less than 1,000 meters.



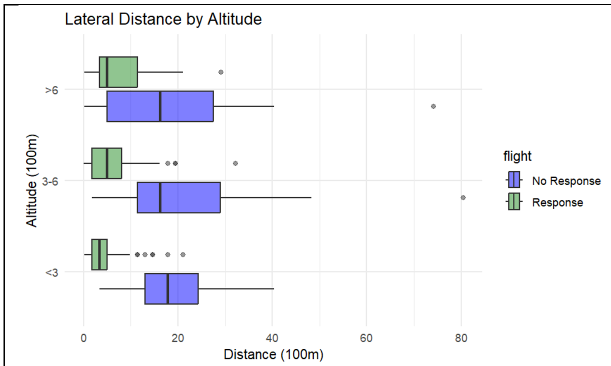
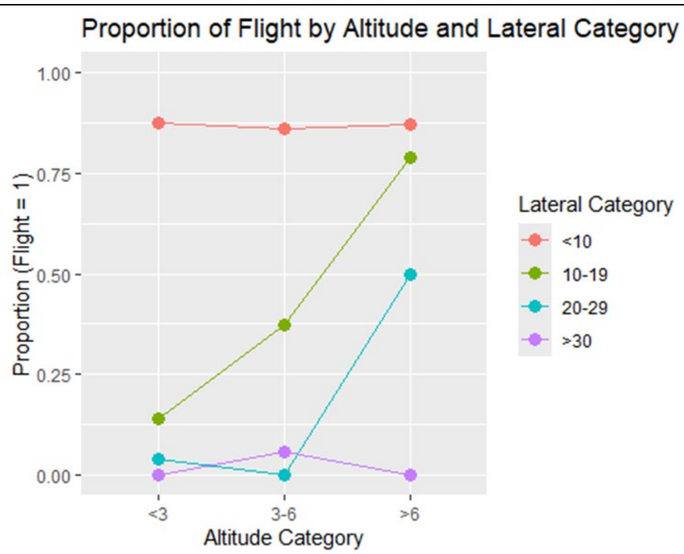
While by nature altitude would be considered a continuous variable, each flight was randomly assigned to one of 9 discrete levels. Most observed flights were under 600 meters.



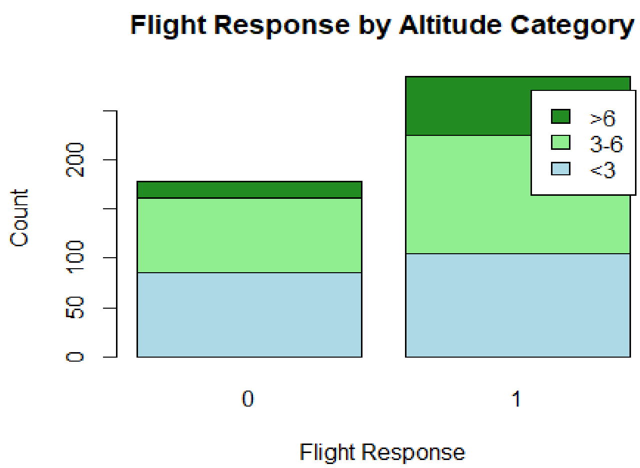
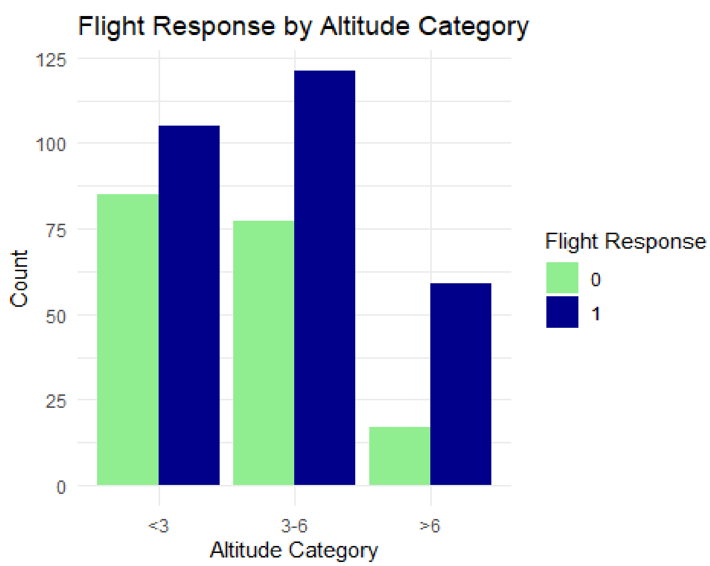
A quick comparison of lateral distance (distance) to responsiveness for each observation shows that closer distances generated a response from the observed flocks.



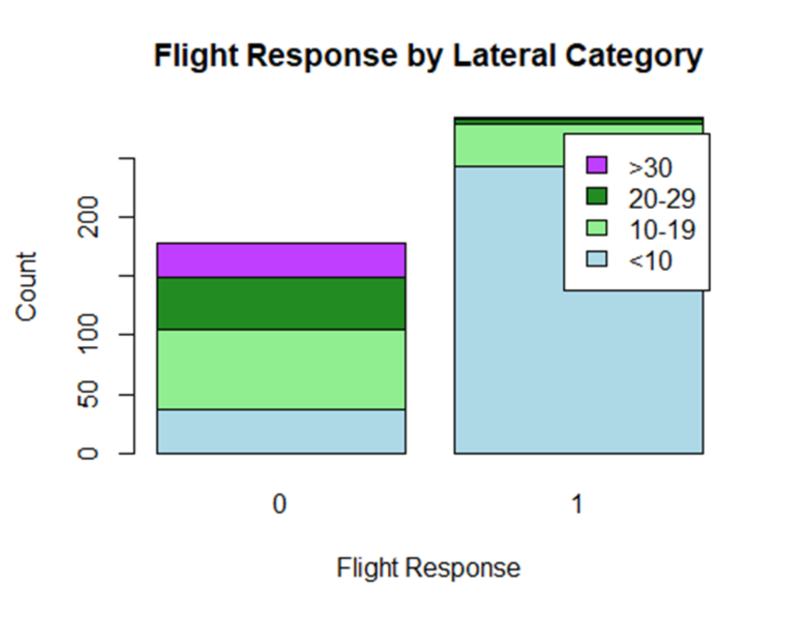
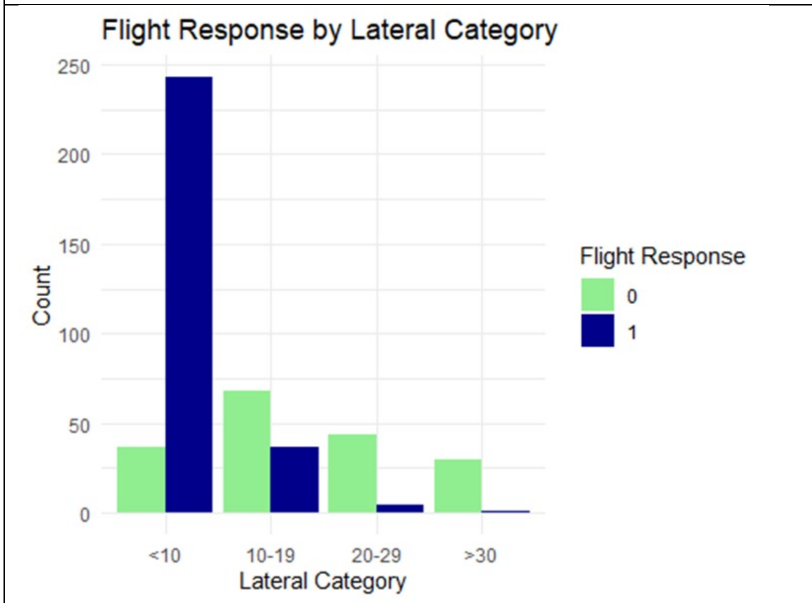
Looking at the data by observation we can see that flocks had responses to flights at all classes of altitude.

	The difference in means between the flight responses at all distances is apparent in this diagram.
	Plotting proportions, altitudes in the shortest lateral distance have a high proportion of flight responses and the line is relatively flat indicating at that close distance, there was not a significant interaction between altitude and lateral distance. An interaction with the two highest brackets for lateral distance, more than 3000m away, appears in the height of 300-600m. This chart illustrates a higher proportion of flight responses.

Exploring Altitude

	Overall, there were more flight responses than non-responses. Both flight responses and non-responses occurred at every altitude category. At the >6 altitude category, the plot indicates that significantly more flights had a response than did not have a response.
	A slightly different visualization illustrates that more flocks had a response than did not have a response at every category of altitude.

Exploring Lateral Distance

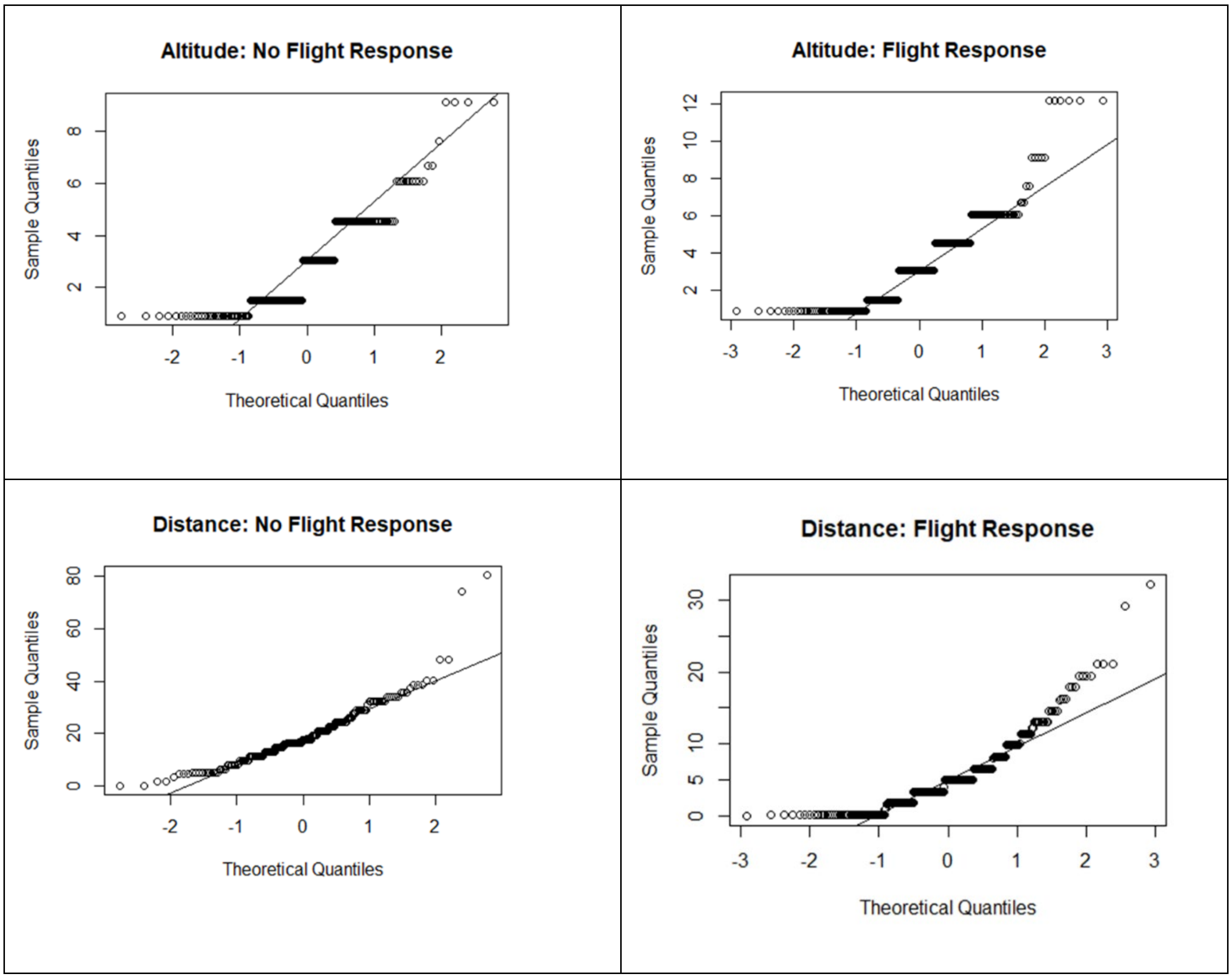
<table>
<tr>
<td>
</td>
<td>There were significantly more flight responses at the closest lateral distance. While non-responses were spread across the data in each bracket of distance, by far the greatest number of responses occurred in the closest bracket.</td>
</tr>
<tr>
<td>
</td>
<td>This plot shows an overwhelmingly higher number of close (<1000 m) helicopter flights causing a flight response. In all other distances, there were more non-responses than responses, with a significant drop off proportionally once the helicopters were more than 2000 m away.</td>
</tr>
</table>

DATA TESTING

Initial exploration suggests that flocks of Pacific brant are not only impacted by the altitude but by the lateral distance of helicopters as well. To further assess the responsiveness of these predictors we can develop statistical models. In the course of that work, understanding whether the data is normally distributed help guide the statistical techniques and tools we'll use to further our analysis.

Evaluating normality

Normal Q-Q plots suggest that data is exhibiting some qualities of non-normality across the various attributes we are evaluating.



Chi-square test: Given evidence of non-normality we used the nonparametric Chi-square test for evaluating the relationship between our variables. The P-value is less than the significance level of 0.05, so we reject the null hypothesis and conclude that there is a statistically significant association between altitude category and flight response.

```
Pearson's Chi-squared test

data:  table(brant_1$altitude_cat, brant_1$flight)
X-squared = 11.477, df = 2, p-value =
0.00322
```

Does flight response have a different mean altitude or lateral distance? To answer this, we determine the difference in the calculated means of our response vs. non-response group without considering categories of altitude or lateral distance. Initial calculations show the mean for altitude is lower in the non-response group, while the mean for lateral distance is significantly smaller for the response group. Determining whether these differences are significant and not due to random error requires additional testing.

Welch's t-test: To verify the difference in mean was significant we used Welch's t-test to test for comparing the means of flocks that responded to the flight compared to those that did not. We performed a t-test for lateral distance.

```
        Welch Two Sample t-test (lateral)

data:  brant_1$lateral by brant_1$flight
t = 15.034, df = 223.16, p-value <
2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 12.11484 15.77001
sample estimates:
mean in group 0 mean in group 1
       19.53436         5.59193
```

The test returned a low p-value and the confidence interval contained 0.

Wilcoxon sign test: We followed the t-test with the Wilcoxon sign test which confirmed there is some impact.

```
Wilcoxon rank sum test with continuity correction (altitude)

data:  altitude by flight
W = 22692, p-value = 0.04121
alternative hypothesis: true location shift is not equal to 0
```

```
Wilcoxon rank sum test with continuity correction (lateral distance)

data:  lateral by flight
W = 45523, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

MODELING

Given the response is categorical, we are preparing to develop linear regression model. While there is evidence of non-normality the sample size was large enough that we think we can iteratively find a linear regression model to represent this population.

Model 1: The first logistic model was a simple model evaluating flight against altitude category. The logistic regression model predicts the probability of a flight response based on the altitude category.

H_o: Altitude category has no effect on flight response

H_a: At least one altitude category impacts flight response

```
Call:
glm(formula = flight ~ altitude_cat, family = binomial, data = brant_1)


Coefficients:
                Estimate Std. Error z value
(Intercept)       0.2113     0.1459   1.448
altitude_cat3-6   0.2407     0.2063   1.167
altitude_cat>6    1.0330     0.3115   3.316
                Pr(>|z|)
(Intercept)     0.147545
altitude_cat3-6 0.243250
altitude_cat>6  0.000914 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)


    Null deviance: 618.81  on 463  degrees of freedom
Residual deviance: 606.71  on 461  degrees of freedom
AIC: 612.71
Number of Fisher Scoring iterations: 4
```

The result of Model 1 shows the p-value for altitude category >6 is 0.0009. The p-value is less than the significance level of 0.05, so we reject the null hypothesis and conclude that helicopters flying at a level of >6 have a significant impact on flight response at the 5% level of significance. The p-values of the <3 and 3-6 altitude categories are greater than 0.05, indicating that helicopters flying at those altitudes do not have a significant effect on flight response.

Using an iterative approach, we ran a few more models similarly evaluating each variable before identifying the one that we thought had the best fit. We expanded the model to include the category of lateral distance, we filtered variables that we identified as insignificant in earlier models and finally added an interaction between altitude and lateral distance. The table below shows a summary of the models, the modifications we made along with the degrees of freedom and AIC scores

Model #	Modification	Df	AIC
Model 2	Incorporated latitude	261	216.17
Model 3	Excluded altitude categories	248	214.17
Model 4	Excluded lateral distance categories	248	214.17

Model 5**:** For our final model, we explored the interaction between altitude and distance. We started from our last model, which was an improvement over the prior models and incorporated an interaction between altitude and lateral distance.

```
Call:
glm(formula = flight ~ altitude_cat + lateral_cat + altitude_cat *
    lateral_cat, family = binomial(link = logit), data = brant_filtered_2)
Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                       1.9459     0.2857   6.811 9.71e-12 ***
altitude_cat>6                   -0.0241     0.5222  -0.046 0.963193
lateral_cat10-19                 -3.7651     0.5247  -7.175 7.21e-13 ***
lateral_cat20-29                 -5.1240     1.0596  -4.836 1.33e-06 ***
altitude_cat>6:lateral_cat10-19   3.1650     0.8849   3.577 0.000348 ***
altitude_cat>6:lateral_cat20-29   3.2022     1.4073   2.275 0.022882 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 326.06  on 251  degrees of freedom
Residual deviance: 191.32  on 246  degrees of freedom
AIC: 203.32
Number of Fisher Scoring iterations: 5
```

In the last model we ran the degrees of freedom decreases as does the AIC score. Most values are significant, the exception is the altitude category for flights more than 600 meters. We did not want to exclude it to evaluate its interaction with the other categories of distance.

CALCULATING PROBABILITY

From our output we can calculate the log odds which represents the probability of success to failure. The log odds can be calculated from the following equation:

log = 1.9459 - .0241(altitude>6) - .37651(lateral 10-19) – 5.1240(lateral 20-29) + 3.1650(altitude>6 * lateral 10-19) + 3.2022 (altitude >6 * lateral 20-29)

To calculate the log odds, each variable in parenthesis is captured as 1 or 0 depending on whether the condition is met; 1 if the condition is met. For example, if a helicopter is flying at an altitude of more than 600 meters and a distance between 1,000 and 1,900 meters the log odds would be calculated as follows:

log = 1.9459 - .0241(1) - .37651(1) – 5.1240(0) + 3.1650(1) + 3.2022 (0)

log = 1.9459 - .0241(1) - .37651(1) + 3.1650(1) = 4.71

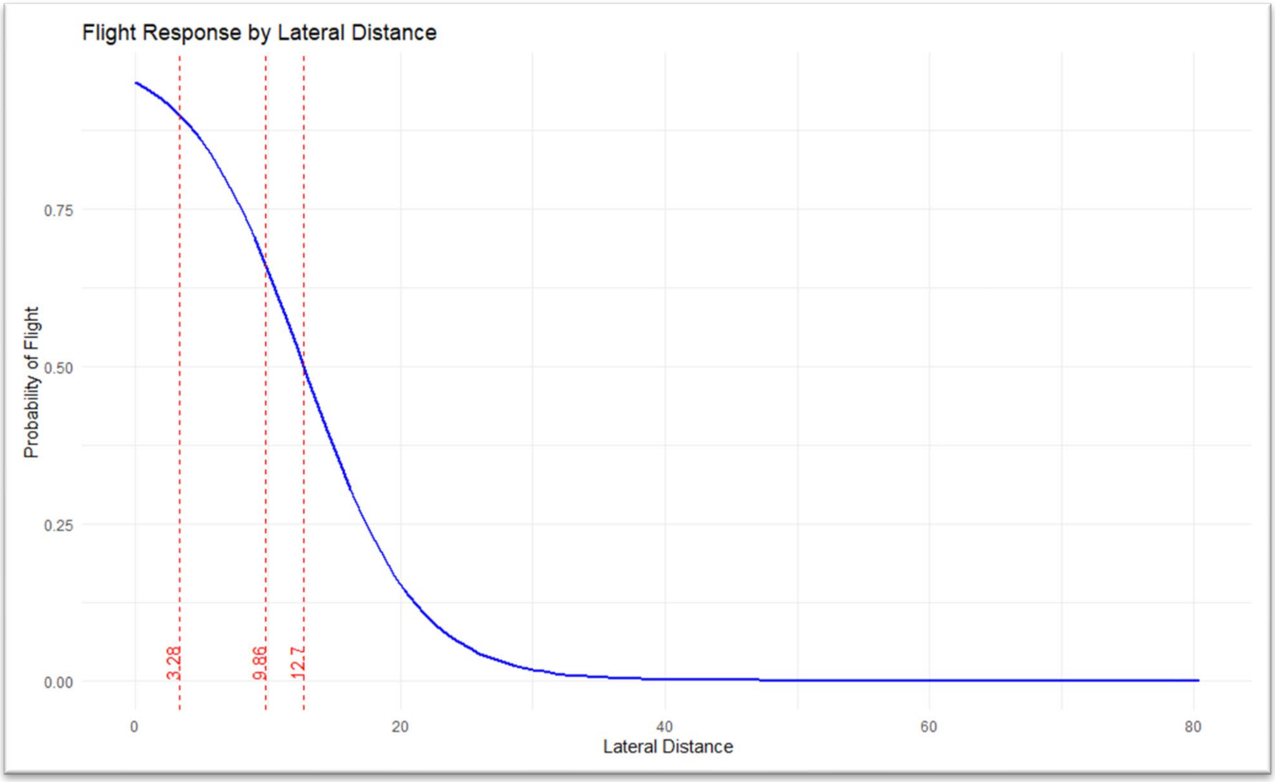
The log odds is used as an input for computing probability of a flight response in the following equation:

P(flight=1)=

P(flight=1)= = .99

At this combination of distance and altitude we can expect a 99% chance of a response to the flight.

The image below converts the logistic model that only considers lateral distance into a probability with line at the distance where there is a probability of .5, .66 and .9 of there being a flight response.



Based on this, a recommended minimal lateral distance of 12,700 m for helicopters landings would ensure less than 50% chance of flight response from the brant.

RECOMMENDATION

Currently the FAA has regulations around altitude but not distance. Based on our findings, we recommend that the FAA retain the regulations around altitude and impose similar restrictions around lateral distance, which appears to have an even more profound effect on the brant.

With only four variables, there were some limitations to the data collected. We felt that the study and analysis could be more robust if the following additional questions or data points were collected as part of the study:

- Was there is a specific time of day the data was collected or was it throughout the day?
- What were the specific responses this study addresses beyond the brant flight from their nesting pattern? Are there other categories of response that could help evaluate the severity?

Regardless, there is clearly an impact from human activity that could be addressed with expanded restrictions.