Comparing False Crawls and Pseudo-Absence Points when Modeling Sea Turtle Nesting on Pensacola Beach

Ashley Ordonez
UWF Proseminar
Advisor: Dr. Seals

Statement of Problem

- A small portion of loggerhead sea turtles nest in Pensacola Beach.
  - Many studies have focused on the loggerhead population elsewhere.
  - Our research attempts to fill in the gaps for the Panhandle area.
- We can identify existing nests but it is difficult to know where sea turtles choose not to nest.
  - Pseudo-absence points.
  - Observed false crawls.
- We have options for how to evaluate locations without nests.
  - How do pseudo absence points compare to false crawl points?
  - How do multiple binary regression models compare to a single multinomial regression model?

Literature Review: Lab History

- J. Iserman (2022): *Beach Morphology Characteristics of Sea Turtle Nesting Sites: A Statistical Analysis*
  - Comparison of beach characteristics within 100 m of a nesting site.
  - Only evaluated present nests.
  - Wilcoxon signed rank showed a difference in foreshore slope and beach slope.

- A. Jensen (2024): *An Exploration of Presence and Pseudo-Absence Data in the Analysis of Loggerhead Sea Turtle Nesting Behavior in the Florida Panhandle*
  - Pseudo-simulation: dataset for 10:1 ratio was resampled to generate different ratios of pseudo-absence points to presence points.
  - Logistic regression used to evaluate bias across ratios.

Literature Review: Lab History

- C. Long (2024): *Modeling Loggerhead Nesting Patterns: How Many Pseudo-Absence Points are Necessary?*
  - Full simulation study for 1:1, 2:1, 5:1, and 10:1 ratios to know true relationships.
  - Results showed bias for beta increased as the ratio increased.

- A. Scamardo (2025): *From the Sea to Statistics: Using Machine Learning to Predict Sea Turtle Nesting Patterns*
  - Compared machine learning methods to examine accuracy in predicting nesting status between presence and pseudo-absence points.
  - Results showed that the random forest approach had higher accuracy, sensitivity, and specificity than the support vector machine approach.

Literature Review: Other Researchers

- H. Afford (2016): *Using multivariate analysis to determine characteristics of sea turtle nest selection along the Florida Panhandle*
    - Examined nest selection of loggerhead sea turtles in Okaloosa county.
    - Pseudo-absence points were used as “no nest” sites.
    - Bayesian logistic regression, classification trees, and random forest analyses were employed to identify environmental predictors of nesting.

- L. Hernandez (2025): *Sea Turtle Nesting on Nourished Beaches With Different Construction Designs: A Case Study in Southeast Florida, USA*
    - Examined nest selection of loggerhead and green sea turtles in South Florida.
    - False crawl points were used as “no nest” sites.
    - Descriptive statistics and zone-based proportions were used to compare, rather than inferential statistics.

Literature Review: Other Researchers

- B. Byrd (2022): *Effects of Environmental and Anthropogenic Factors on Loggerhead Sea Turtle (Caretta Caretta) False Crawl Rates on Jekyll Island, GA*
  - Examined nest selection of loggerhead sea turtles on Jekyll Island, GA.
  - False crawl points were used as “no nest” sites.
  - Bayesian logistic regression was employed to identify predictors of nesting.

- S. Manestar (2023): *Using Unmanned Aerial Vehicle (UAV) Surveys and Traditional Methods to Examine Influences on Loggerhead Sea Turtle (Caretta caretta) Nest Site Selection*
  - Examined nest selection of loggerhead sea turtles in Boca Raton, FL, considering structural impediments.
  - False crawl points were used as “no nest” sites.
  - Logistic regression was employed to identify predictors of nesting.

Current Study: Design

- Overall research question: **how do pseudo absence points compare to false crawl points?**

- All data was observed on Pensacola Beach
  - P: present nests – nests were observed by state employees
  - PA: pseudo-absence points – using ArcGIS, observations were randomly selected from the areas *without* observed nests
  - FC: false crawl points – observed turtle tracks in the sand with a turn around point.

Current Study: Data

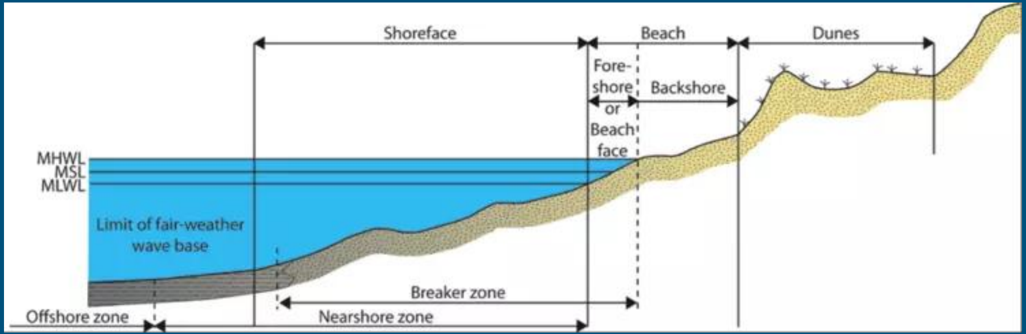


Image credit: M. Kulp

- Outcome:
    - Nest type (nominal):

      presence, pseudo-absence, false crawl
- Predictors:
    - Foreshore slope (continuous): slope (degrees) of the profile from the mean low-water line (MLWL) to the mean high-water line (MHWL).
    - Beach slope (continuous): slope (degrees) of the profile from the mean high-water line (MHWL) to the potential line of vegetation.

Binary Logistic Regression

Binary Logistic Regression is used to predict the probability of an outcome when there are only two possible outcomes.

It is modeled in the form:

$$\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

Where $\pi = \mathrm{P}[Y = 1]$ represents the probability of the outcome of the event.

Multinomial Logistic Regression

Multinomial Logistic Regression is used to predict the probability of an outcome occurring when there are three or more unordered outcomes.

It is modeled in the form $$\ln\left(\frac{\hat{\pi}_j}{\hat{\pi}_{\text{ref}}}\right) = \hat{\beta}_{0j} + \hat{\beta}_{1j}x_1 + \ldots + \hat{\beta}_{kj}x_k$$

where *j* is the response category, *ref* represents the reference category, and there are *k* predictors and the probability of the outcome of the events are given by

$$\pi = \mathrm{P}[Y = 1]$$

This creates *c* - 1 models where *c* is the number of outcomes.

Interpreting the Models

The initial models for Binary Logistic Regression and Multinomial Logistic Regression provide the log odds for each predictor.

$$\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

$$\ln\left(\frac{\hat{\pi}_j}{\hat{\pi}_{\text{ref}}}\right) = \hat{\beta}_{0j} + \hat{\beta}_{1j} x_1 + \ldots + \hat{\beta}_{kj} x_k$$

Since we are modeling the natural log of our odds, we need to “undo” the log to interpret the models. Thus, we exponentiate the outcomes to find the probability of the outcome occurring.

- This is called the odds ratio.

Interpreting the Models

Direction of relationship is indicated by value:

- β > 0 or OR > 1 indicates an increase
- β = 0 or OR = 1 indicates no change
- β < 0 or 0 ≤ OR < 1 indicates a decrease

$$\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

$$\ln\left(\frac{\hat{\pi}_j}{\hat{\pi}_{\text{ref}}}\right) = \hat{\beta}_{0j} + \hat{\beta}_{1j} x_1 + \ldots + \hat{\beta}_{kj} x_k$$

Beta interpretation:

- For a 1 unit increase in x, the log odds of the outcome occurring will be increase by β.

Odds ratio interpretation:

- For a 1 unit increase in x, the odds of the outcome occurring will be multiplied by e^{β}.

Inference: Confidence Intervals

Confidence interval for β:

$$\hat{\beta}_i \pm Z_{\frac{\alpha}{2}} * SE_{\beta_i} = (LL, UL)$$

Confidence interval for OR:

$$(e^{LL}, e^{UL})$$

Inference: Hypothesis Testing

To test for the significance of our predictors, we use the Wald Z test.

Hypotheses: $H_0 : \beta_i = 0$ $H_a : \beta_i \neq 0$

We test this significance by calculating the Z-score using the following formula: $Z^* = \left(\frac{\hat{\beta}_i}{s(\beta_i)} \right)$

For the above formula, Z^* is the test statistic, β_i is the slope of the i predictor, and $s(\beta_i)$ is the standard deviation of the i predictor. Once we determine the confidence level α, we reject if $p < \alpha$.

Current Study: Models Considered

Binomial logistic regression models under consideration:

- Modeling P/PA
$$\ln\left(\frac{\hat{\pi}_{\mathrm{P}}}{1-\hat{\pi}_{\mathrm{P}}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$
- Modeling P/FC
$$\ln\left(\frac{\hat{\pi}_{\mathrm{P}}}{1-\hat{\pi}_{\mathrm{P}}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

Multinomial logistic regression models under consideration:

- Modeling P/PA
$$\ln\left(\frac{\hat{\pi}_{\mathrm{P}}}{\hat{\pi}_{\mathrm{PA}}}\right) = \hat{\beta}_{0j} + \hat{\beta}_{1j} x_1 + \ldots + \hat{\beta}_{kj} x_k$$
- Modeling P/FC
$$\ln\left(\frac{\hat{\pi}_{\mathrm{P}}}{\hat{\pi}_{\mathrm{FC}}}\right) = \hat{\beta}_{0j} + \hat{\beta}_{1j} x_1 + \ldots + \hat{\beta}_{kj} x_k$$

Assessing Model Performance

- Accuracy
  - (TP + TN) / (TP + TN + FN + FP)
- Sensitivity
  - TP / (TP + FN)
- Specificity
  - TN / (FP + TN)
- Positive Predictive Value
  - TP / (TP + FP)
- Negative Predictive Value
  - TN / (FN + TN)

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Current Study: Overall Goal

- How different are analyses when using PA instead of FC?
  - PA:
    - Positive: because these are randomly generated, we can collect an infinite number.
    - Negative: We do not know if turtles actually considered that spot for nesting.
  - FC:
    - Positive: We do know that turtles considered the spot for nesting.
    - Negative: This data collection relies on government employees observing the tracks.

- Goal: quantify differences in model results / predictions

Results: Binary Logistic

- Modeling P/PA

$$\ln\left(\frac{\hat{\pi}_P}{1-\hat{\pi}_P}\right) = 0.308 + 0.022 * BeachSlope + 0.013 * ForeshoreSlope$$

- Modeling P/FC

$$\ln\left(\frac{\hat{\pi}_P}{1-\hat{\pi}_P}\right) = 0.88 - 0.049 * BeachSlope + 0.011 * ForeshoreSlope$$

Results: Multinomial Logistic

- Modeling P/PA

$$\ln\left(\frac{\hat{\pi}_{\mathrm{P}}}{\hat{\pi}_{\mathrm{PA}}}\right) = -0.759 + 0.088 * BeachSlope + 0.051 * ForeshoreSlope$$

- Modeling P/FC

$$\ln\left(\frac{\hat{\pi}_{\mathrm{P}}}{\hat{\pi}_{\mathrm{FC}}}\right) = 0.89 - 0.048 * BeachSlope + 0.009 * ForeshoreSlope$$

Results: Model Comparisons

When comparing P to PA:

Modeling Approach	**Predictor**	**OR (95% CI)**	**p-value**
Binary	Foreshore Slope	1.01 (0.98 - 1.05)	0.460
	Beach Slope	1.02 (0.99 - 1.05)	0.154
Multinomial	Foreshore Slope	1.05 (0.92 - 1.21)	0.467
	Beach Slope	1.09 (0.97 - 1.23)	0.162

Results: Model Comparisons

When comparing P to FC:

Modeling Approach	Predictor	OR (95% CI)	p-value
Binary	Foreshore Slope	1.01 (0.85 - 1.20)	0.899
	Beach Slope	0.95 (0.83 - 1.10)	0.487
Multinomial	Foreshore Slope	1.01 (0.85 - 1.20)	0.914
	Beach Slope	0.95 (0.83 - 1.09)	0.490

Results: Classification

Binary logistic:

	True	
Pred	P	PA
P	38	27
PA	41	51

	True	
Pred	P	FC
P	79	38
FC	0	0

Multinomial logistic:

	True		
Pred	P	PA	FC
P	38	27	20
PA	41	51	18
FC	0	0	0

Results: Classification

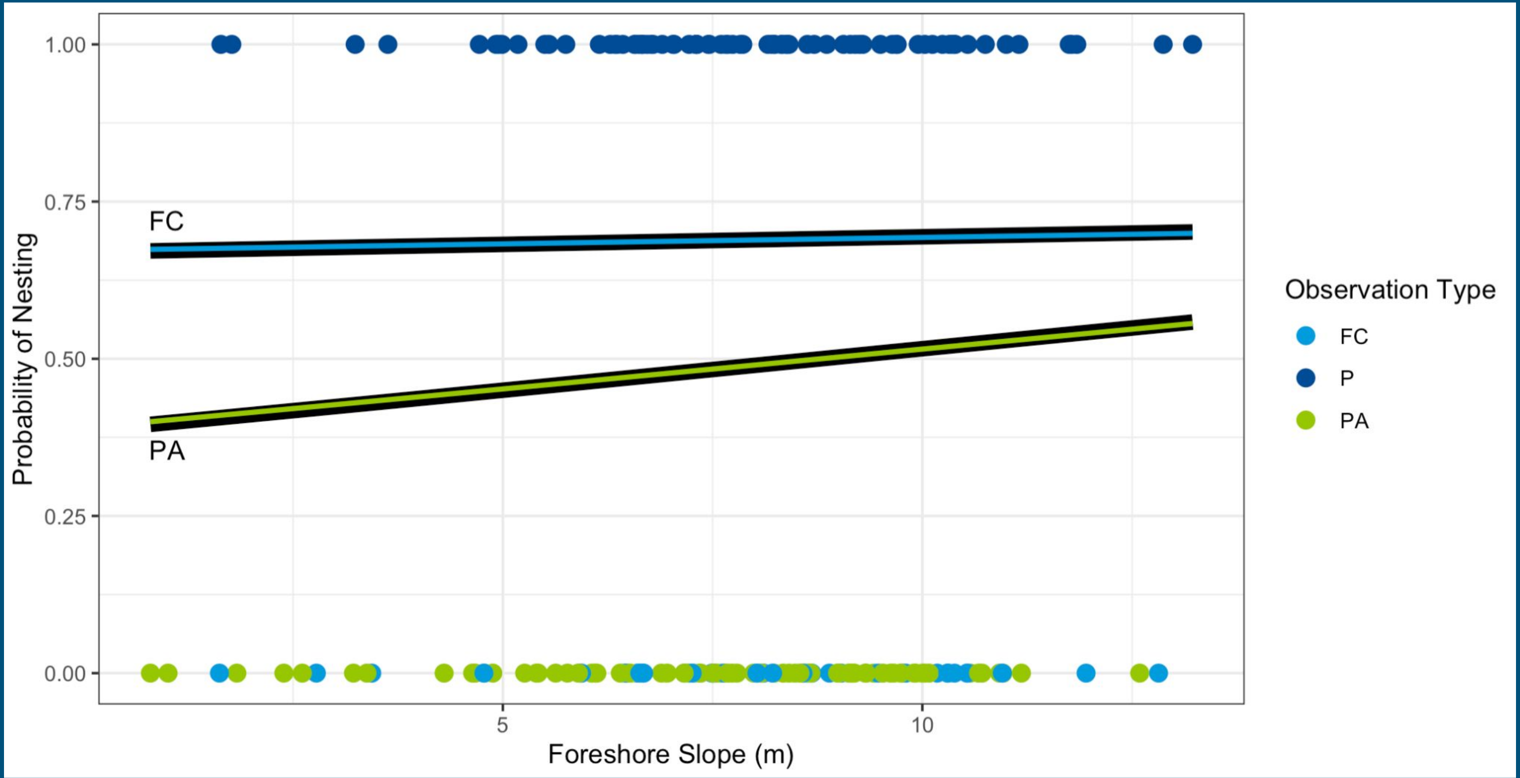
Binary logistic (P vs PA) – *model struggles to distinguish P & PA* ● Accuracy ≈ 57% ● Sensitivity (P) = 48% ● Specificity (PA) = 65% ● Positive Predictive Value = 58% ● Negative Predictive Value = 55%	Binary logistic (P vs FC) – *model never predicts FC* ● Accuracy ≈ 68% ● Sensitivity (P) = 100%, ● Specificity (FC) = 0% ● Positive Predictive Value = 68% ● Negative Predictive Value = 0%

Results: Classification

- Multinomial logistic (P vs PA vs FC) – *improves P vs PA prediction, but fails to predict FC*
  - Accuracy ≈ 46%

Sensitivity: • P = 45%, • PA = 46% • FC = 0%	Specificity: • P = 63%, • PA = 68% • FC = 81%
Positive Predictive Value: • P = 48% • PA = 65% • FC = 0%	Negative Predictive Value: • P = 59% • PA = 50% • FC = 100%

Results: Model Visualization



Conclusions

- There is not a statistical relationship between sea turtle nesting on Pensacola Beach and either foreshore slope or beach slope.
  - Nesting habits may be driven by other factors unavailable to us at this time.
- There is little, if any, difference between modeling approaches.
  - Regression coefficients are different, but not in a meaningful way.
  - The data's story does not change.
- Our models struggle to discern between FC and PA observations.

Limitations

- Both P and FC are human observed, while PA is randomly generated.
  - The P are observed nests.
  - The FC are observed tracks where a turtle turns around.
  - PA points are generated using ArcGIS.
- We do not know the true relationships between nesting status and beach slope or foreshore slope.
- A smaller population of sea turtles nests in the Pensacola area – larger datasets can be found elsewhere.

Suggestions of Further Study

- We do not know the true relationships between nesting status and beach slope or foreshore slope.
  - Simulation study to define the relationship, then explore modeling errors.
- A smaller population of sea turtles nests in the Pensacola area – larger data can be found elsewhere.
  - Expand data collection to areas outside of Pensacola to expand on comparing FC and PA points.
- Directly compare FC and PA points.

References

- Agresti, Alan. 2018. *An Introduction to Categorical Data Analysis*. Third ed. Hoboken, NJ: Wiley.
- Anon. 2016. “Nearshore, Beaches, and Dunes.” *InTeGrate*. Retrieved November 18, 2025 (https://serc.carleton.edu/166724).
- Anon. 2025. “Confusion Matrix.” *Wikipedia*. Retrieved November 18, 2025 (https://en.wikipedia.org/wiki/Confusion_matrix).

Thank you