

# NRES 776 Lab 9

GLM with single predictor

Sunny Tseng

# Our schedule today

- Discussion paper presentation (50 min)
- Binomial regression with single continuous  $x$  (15 min)
- Lab intro - Poisson regression with single predictor variable (10 min)
- Lab intro - Binomial regression with single predictor variable (10 min)
- Practice (60 min)

# Discussion paper presentation

- Olivia Wilson
- Tina Watters
- Brett Gandy

# Binomial regression

# Research question (1 continuous x)

A survey was done on 50 islands for the incidence of a bird species Grasshopper Warbler. Researchers want to know whether the incidence is related to the area and/or the isolation level of the islands.

## Use long format as input

```
# A tibble: 50 × 3
  incidence area isolation
  <fct>     <dbl> <fct>
1 1         7.93 low
2 0         1.92 high
3 1         2.04 median
4 0         4.78 median
5 0         1.54 median
6 1         7.37 low
7 1         8.60 low
8 0         2.42 high
9 1         6.40 median
10 1        7.20 median
# i 40 more rows
```

## Take a look at the relationship

```
# A tibble: 3 × 3
  proportion total mean_area
  <dbl> <int>     <dbl>
1      1      15      6.03
2  0.636      22      3.93
3      0      13      3.01
```

# Model formulation

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{1i}$$

```
1 incidence_binom <- glm(formula = incidence ~ area,  
2                         data = data_long,  
3                         family = "binomial")  
4 incidence_binom %>% summary
```

Call:

```
glm(formula = incidence ~ area, family = "binomial", data = data_long)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-2.1554	0.7545	-2.857	0.004278	**
area	0.6272	0.1861	3.370	0.000753	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 68.029 on 49 degrees of freedom  
Residual deviance: 50.172 on 48 degrees of freedom  
AIC: 54.172

Number of Fisher Scoring iterations: 5

# Coef. interpretation

- For island with area as 0 (baseline):

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0$$

$$\frac{p_i}{1 - p_i} = \text{Odd}(p_i) = \exp(\beta_0) = \exp(-2.15) = 0.11$$

- The odd for the intercept is not often interpreted by itself.

# Coef. interpretation (con'd)

- For island with non-0 area:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{1i}$$

$$\text{logit}(p'_i) = \beta_0 + \beta_1 (x_{1i} + 1)$$

$$\text{logit}(p'_i) - \text{logit}(p_i) = \beta_1 = \log\left(\frac{\text{Odd}(p'_i)}{\text{Odd}(p_i)}\right)$$

$$\frac{\text{Odd}(p'_i)}{\text{Odd}(p_i)} = \exp(\beta_1) = \exp(0.62) = 1.87$$

- For one unit increase in **area**, the odds of the bird species being present increase by a factor of 1.87.



# Output interpretation

## Odd ratio (OR)

- $OR = 1$ , no difference between groups
- $OR < 1$ , treatment decreases odds
- $OR > 1$ , treatment increases odds

## R output

- **Intercept (-2.15)**: The odd of the bird being present on 0 area island is  $\exp(-2.15)$
- **area (0.62)**: For one unit increase in **area**, the odds of the bird species being present increase by a factor of  $\exp(0.62)$
- **Dispersion parameter (1)**: wonderful. Need to consider other methods if dispersion larger than 1 (over-dispersion) or smaller than 1 (under-dispersion)
- **AIC (54.172)**: Can be used to compare the goodness of fit between models

# Model goodness of fit: Likelihood ratio test

```
1 incidence_binom_null <- glm(formula = incidence ~ 1,  
2                             data = data_long,  
3                             family = "binomial")
```

$H_0$ : The model performance is the same as a null model (making predictions by chance)

$H_1$ : The model performance is significantly different comparing to a null model

- Use `lrtest()` function in the `lmtest` package

```
1 lrtest(incidence_binom, incidence_binom_null)
```

Likelihood ratio test

Model 1: incidence ~ area

Model 2: incidence ~ 1

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	2	-25.086			
2	1	-34.015	-1	17.857	2.382e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Model prediction

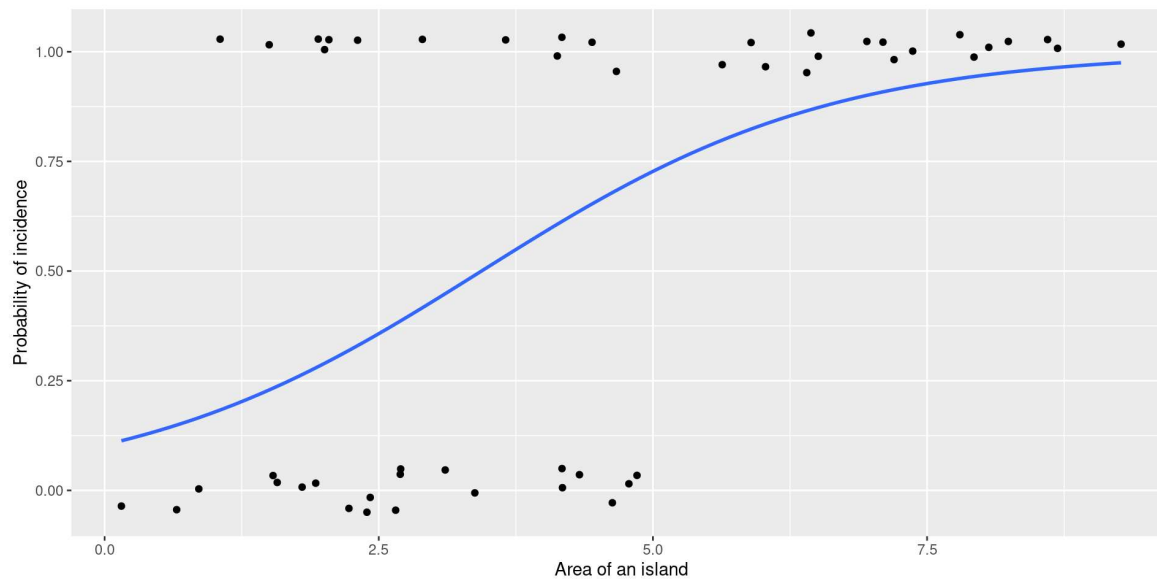
- Use `predict()` and specify `type = "response"` to get back transformed  $y_i$

```
1 incidence_p <- data_long %>%  
2   mutate(incidence_p = predict(incidence_binom, type = "response"))  
3 incidence_p
```

```
# A tibble: 50 × 4  
  incidence area isolation incidence_p  
  <dbl>    <dbl> <fct>      <dbl>  
1 1      7.93 low      0.944  
2 0      1.92 high     0.279  
3 1      2.04 median    0.295  
4 0      4.78 median    0.699  
5 0      1.54 median    0.233  
6 1      7.37 low      0.922  
7 1      8.60 low      0.962  
8 0      2.42 high     0.346  
9 1      6.40 median    0.865  
10 1     7.20 median    0.914  
# i 40 more rows
```

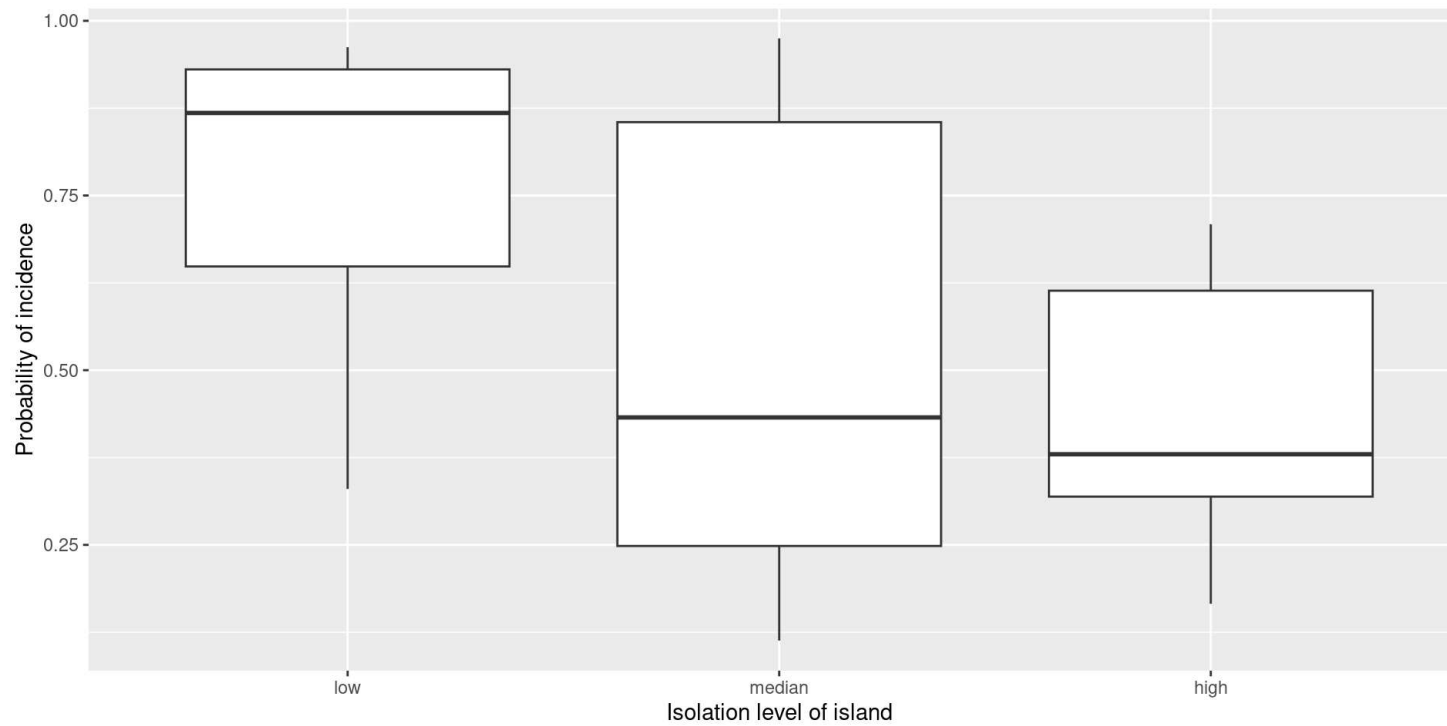
# Model visualization

```
1 data_long %>%  
2   mutate(incidence = incidence %>% as.numeric() - 1) %>%  
3   ggplot(aes(x = area, y = incidence), data = .) +  
4   geom_jitter(width = 0, height = 0.05) +  
5   geom_smooth(method = "glm",  
6               method.args = list(family = "binomial"),  
7               se = FALSE) +  
8   labs(y = "Probability of incidence", x = "Area of an island")
```



# Model visualization

```
1 ggplot(aes(x = isolation, y = incidence_p),  
2       data = incidence_p) +  
3   geom_boxplot() +  
4   labs(y = "Probability of incidence", x = "Isolation level of island")
```



## Lab 9 - GLM with single predictor

# Get the lab assignment file

1. Download the file on Moodle [Lab 9 - assignment file](#). You should get a [.Rmd](#) file after downloading
2. Get the two data set, one crab data set and on titanic data set
3. Create a [R project](#) folder, and put [.Rmd](#), two your data sets in the same folder

## Run R markdown

1. Open the R project, open the [.Rmd](#) file
2. Install and load [rmarkdown](#) package

```
1 install.packages("rmarkdown")  
2 library(rmarkdown)
```

3. Click [knit](#) and you should get a [html](#) output
4. You are all set!

# Submission

**Due next Wednesday 23:59 pm, Moodle**

- File name and type (NRES776\_*firstname\_lastname*\_lab\_9.html)



