Discrete Response Model Lecture 5

Models for Count Response, Discrete Response Model Evaluation, and Model Selection

datascience@berkeley

Parameter Estimation and Inference

Model Interpretation

- We are no longer modeling the log-odds anymore!
- We will not use odds ratios to interpret the effect of

Consider a model with one explanatory variable again:

$$\mu(x) = \exp(\beta_0 + \beta_1 x)$$

where I am using "$\mu(x)$" here to emphasize we are evaluating the model ατ particular numerical value of x. The model evaluated at a c-unit increase in the explanatory variable is

$$\boxed{\mu(x+c)} = \boxed{\exp(\beta_0 + \beta_1 (x + c))}$$

Model Interpretation

- Consider a Poisson regression model with one explanatory variable $\mu(x) = exp(\beta_0 + \beta_1 x)$.
- Increase the explanatory variable by c units, and the result is $\mu(x+c) = exp(\beta_0 + \beta_1(x+c)) = \mu(x)exp(c\beta_1))$.
- The ratio of the means at $x + c$ and x is

$$\boxed{\frac{\mu(x+c)}{\mu(x)} = \frac{exp(\beta_0 + \beta_1(x+c))}{exp(\beta_0 + \beta_1 x)} = exp(c\beta_1)}$$

- The interpretation is that "the percentage change in the mean response resulting from a c unit change in x is $100(e^{c\beta_1} - 1)$"
- As an example, if $exp(c\beta_1) = 1.1$, then it means that the percentage change in the mean response resulting from a c unit change in x is 10%
- Note that this interpretation is not dependent on the original value of x

Model Result Interpretation

```
> summary(crab)
     Color           Spine           Width          Weight           Sat        
 Min.   :1.000   Min.   :1.000   Min.   :21.0   Min.   :1.200   Min.   : 0.000  
 1st Qu.:2.000   1st Qu.:2.000   1st Qu.:24.9   1st Qu.:2.000   1st Qu.: 0.000  
 Median :2.000   Median :3.000   Median :26.1   Median :2.350   Median : 2.000  
 Mean   :2.439   Mean   :2.486   Mean   :26.3   Mean   :2.437   Mean   : 2.919  
 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:27.7   3rd Qu.:2.850   3rd Qu.: 5.000  
 Max.   :4.000   Max.   :3.000   Max.   :33.5   Max.   :5.200   Max.   :15.000  
```

```
Call:
glm(formula = Sat ~ Width, family = poisson(link = log), data = crab)
```

The expected number of satellites when the shell width is 23:

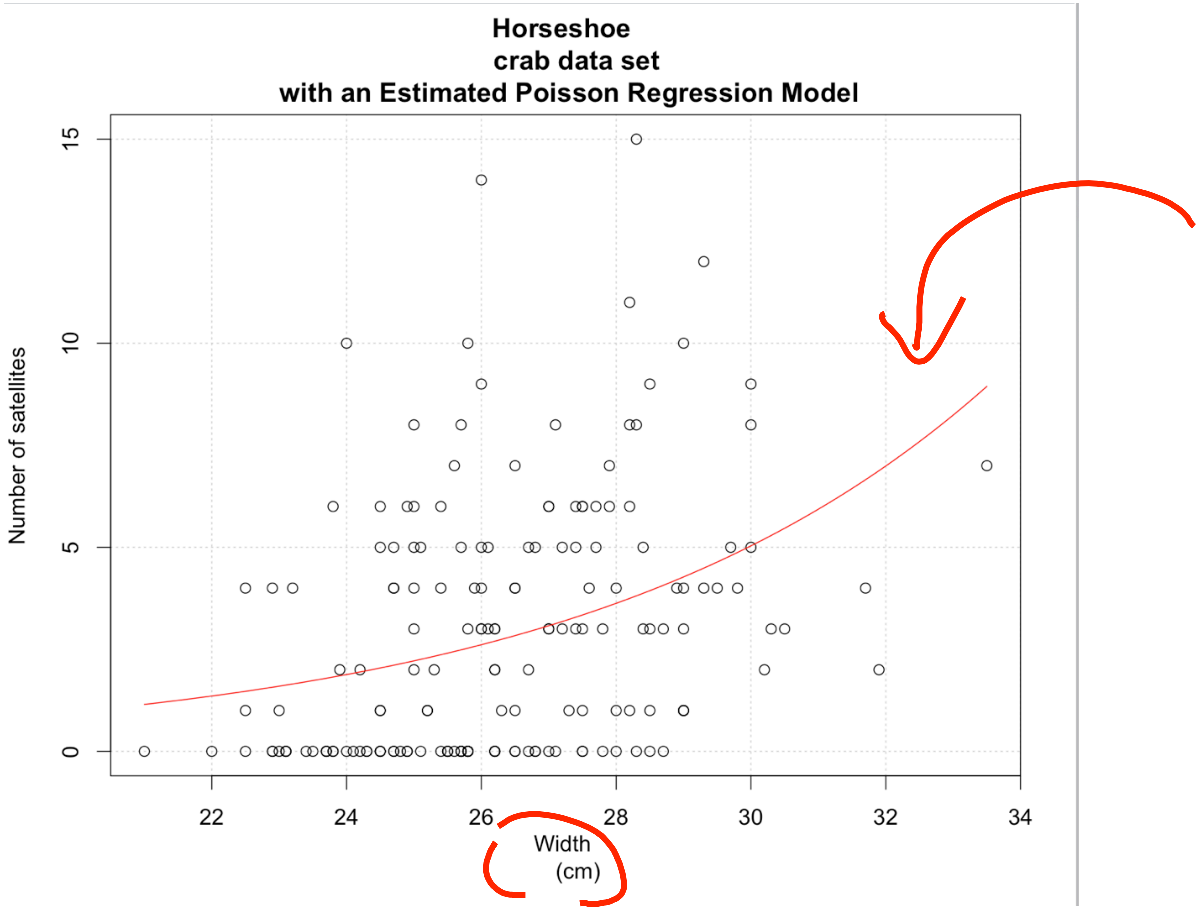
$$\hat{\mu} = \exp(-3.3048 + 0.1640 \times 23) = 1.60$$

- The positive slope indicates that the number of satellite increases with shell width
- So, a 1-unit increase in shell width leads to an 17.8% estimated increase in the number satellites

Example

When there is only one explanatory variable in the model, we can easily examine the estimated model through a plot.

When there are more than one explanatory variables, we will have to make the plot conditional on specific values on all the other variables in the model.



Example

```r
#Function to find confidence interval
ci.mu<-function(newdata, mod.fit.obj, alpha) {
    lin.pred.hat<-predict(object = mod.fit.obj, newdata =
      newdata, type = "link", se = TRUE)
    lower<-exp(lin.pred.hat$fit - qnorm(1 - alpha/2) *
      lin.pred.hat$se)
    upper<-exp(lin.pred.hat$fit + qnorm(1 - alpha/2) *
      lin.pred.hat$se)
    list(lower = lower, upper = upper)
  }
```

```
Browse[1]> ci.mu(newdata = data.frame(Width = 23), mod.fit.obj =
+     mod.fit, alpha = 0.05)
$lower
       1
1.332135

$upper
       1
1.915114
```

Profile Likelihood Ratio Confidence Interval

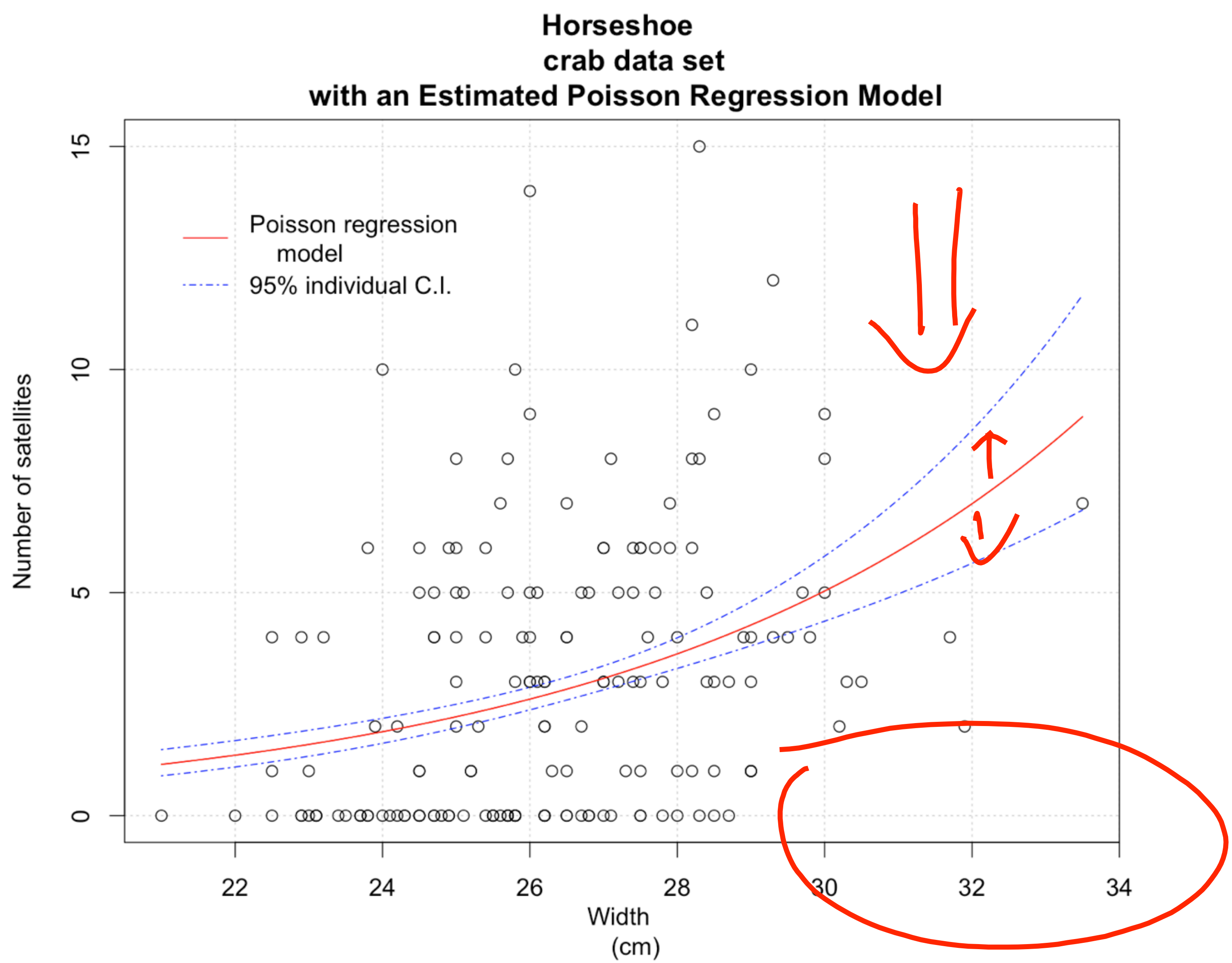
```
library(mcprofile)
linear.combo<-mcprofile(object = mod.fit, CM = K)
#CI for beta_0 + beta_1 * x
ci.logmu.profile<-confint(object = linear.combo, level = 0.95)
ci.logmu.profile
```

```
   mcprofile - Confidence Intervals

level:              0.95
adjustment:         single-step

   Estimate lower upper
C1    0.468 0.284 0.647
```

```
> ci.logmu.profile$confint
      lower     upper
1 0.2841521 0.6471587
> exp(ci.logmu.profile)

   mcprofile - Confidence Intervals

level:              0.95
adjustment:         single-step

   Estimate lower upper
C1      1.6  1.33  1.91
```

The 95% interval is **1.33 < μ < 1.91**, which is quite similar to the Wald interval.

Estimated Model and Confidence Bands



Binning an Explanatory Variable

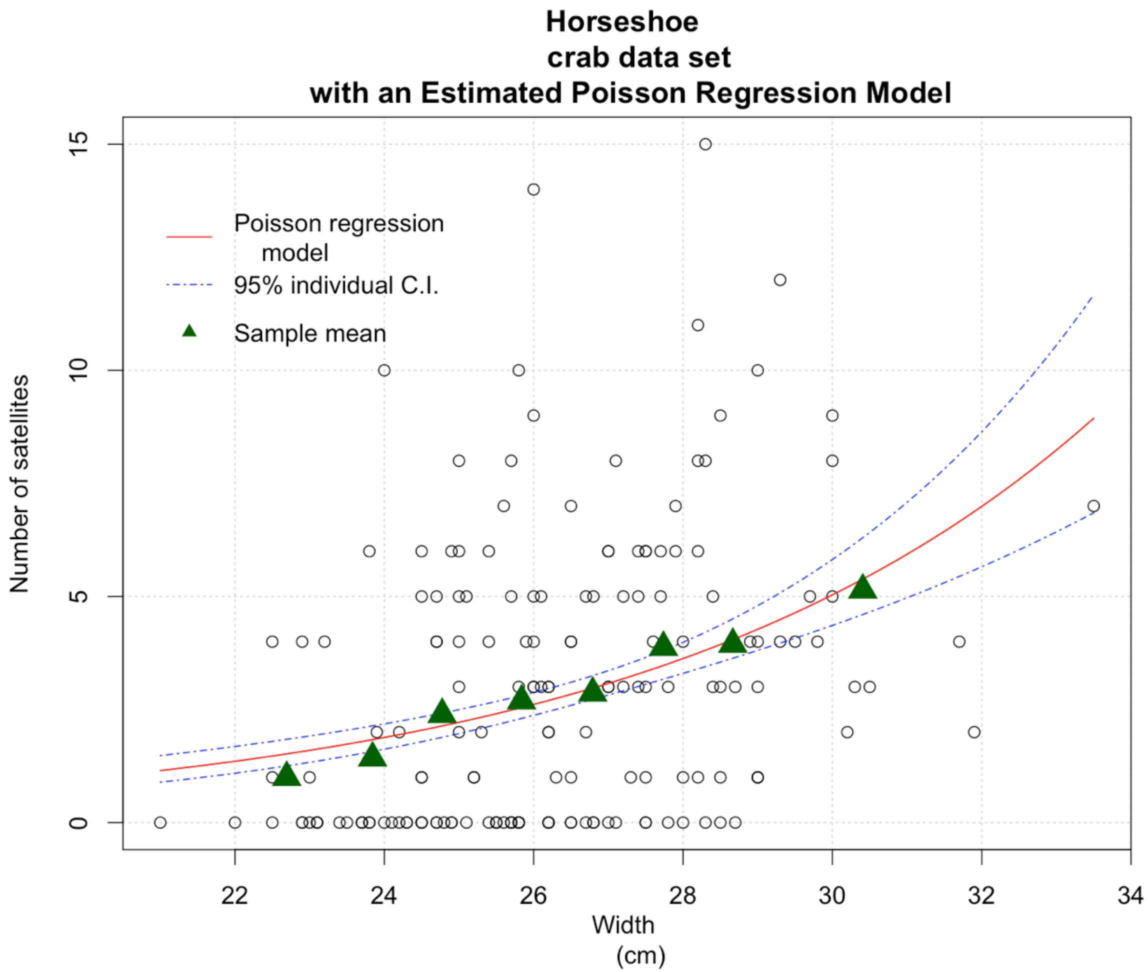
- The data show somewhat of an upward trend. The model captures this through displaying similar qualities.

- You may be alarmed by the number of plotting points far from the estimated model. However, remember that the model is **trying to estimate the "average" number of satellites given the width.**
- We can examine this more closely where we add the average number of satellites for a "width group."

```
Browse[1]> groups<-ifelse(test = crab$Width<23.25, yes = 1, no =
+         ifelse(test = crab$Width<24.25, yes = 2, no =
+         ifelse(test = crab$Width<25.25, yes = 3, no =
+         ifelse(test = crab$Width<26.25, yes = 4, no =
+         ifelse(test = crab$Width<27.25, yes = 5, no =
+         ifelse(test = crab$Width<28.25, yes = 6, no =
+         ifelse(test = crab$Width<29.25, yes = 7, no = 8)))))))
Browse[1]> crab.group<-data.frame(crab,groups)
Browse[1]> head(crab.group)
  Color Spine Width Weight Sat groups
1     2     3  28.3   3.05   8      7
2     3     3  26.0   2.60   4      4
3     3     3  25.6   2.15   0      4
4     4     2  21.0   1.85   0      1
5     2     3  29.0   3.00   1      7
6     1     2  25.0   2.30   3      3
```

```
Browse[1]> ybar<-aggregate(formula = Sat ~ groups, data = crab, FUN = mean)
Browse[1]> xbar<-aggregate(formula = Width ~ groups, data = crab, FUN = mean)
Browse[1]> data.frame(ybar, xbar$Width)
  groups      Sat xbar.Width
1      1 1.000000   22.69286
2      2 1.428571   23.84286
3      3 2.392857   24.77500
4      4 2.692308   25.83846
5      5 2.863636   26.79091
6      6 3.875000   27.73750
7      7 3.944444   28.66667
8      8 5.142857   30.40714
```

Comparing the model mean with the the Average” number of satellites given the width



Notice how the red line goes through the middle of the green triangles (group means).

Final Recap

- This interpretation is not dependent on the original value of x!
- Choose a value of c appropriate for the data.
- The estimate of $100(e^{c\beta_1} - 1)$ is $\cancel{100(e^{c\hat{\beta}_1} - 1)}$.
- Wald and LR confidence intervals can be found using the usual methods.
- If there is more than one explanatory variable in the model, the same result holds but need to included the "conditional" interpretation.
- If there are interactions or transformations of explanatory variables or categorical explanatory variables, similar types of adjustments need to be made as we introduced in previous week.
- The students are referred to the textbook for more information about Poisson regression models.

Berkeley SCHOOL OF INFORMATION