BAT 3302 - Poisson Regression

Malaika Tiglao - An Vu Nguyen Dieu - Thomas Pena

5/8/2022

Introduction to Poisson Regression:

- Poisson regression is a generalized linear model form of regression analysis used to model count data.
    - *Count data: a type of data in which the observations can take only the counting numbers, non-negative integer values (0, 1, 2, 3,…).These integers arise from counting rather than ranking.*
- Poisson regression model: $log(\lambda_i) = \beta_0 + \beta_1 x_i$

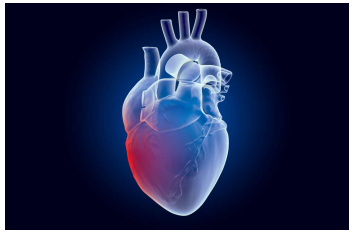
Poisson Regression Assumptions

- Poisson Response: The response variable is a count per unit of time or space, described by a Poisson distribution.
- Independence: The observations must be independent of one another.
- Mean = Variance: By definition, the mean of a Poisson random variable must be equal to its variance.

- Linearity: The log of the mean rate, $log(\lambda)$, must be a linear function of **x.**

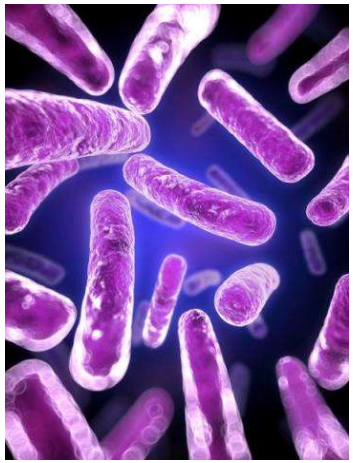
When Poisson Regression should be used:

- Poisson regression, also known as a log-linear model, should be used when the outcome variable is count data.
- Examples:
  - *how many heart attacks or strokes one's had.*



  - *how many days in the past month one's used [insert your favorite illicit substance here].*

- *how many days from outbreak until infection.*



Two Sample Scenarios with A Snip of Data Structure (1):

- Sample Scenario 1: Predict the number of awards earned by a student in high school based on the program they enrolled in and their math final exam score.

```
##     id num_awards       prog math
## 1   45          0 Vocational   41
## 2  108          0    General   41
## 3   15          0 Vocational   44
## 4   67          0 Vocational   42
## 5  153          0 Vocational   40
## 6   51          0    General   42
## 7  164          0 Vocational   46
## 8  133          0 Vocational   40
## 9    2          0 Vocational   33
## 10  53          0 Vocational   46
## 11   1          0 Vocational   40
## 12 128          0   Academic   38
```

Two Sample Scenarios with A Snip of Data Structure (2):

- Sample Scenario 2: Predict the number of warp breaks per loom based on the effect of wool type (A or B) and tension (low, medium, or high).

```
##    breaks wool tension
## 1      26    A       L
## 2      30    A       L
## 3      54    A       L
## 4      25    A       L
## 5      70    A       L
## 6      52    A       L
## 7      51    A       L
## 8      26    A       L
## 9      67    A       L
## 10     18    A       M
## 11     21    A       M
## 12     29    A       M
```