Modelling Count Variables

1. What is Poisson regression used to model. State any assumptions that must be checked before it can be used as an analysis.

2. The R Code output given below is used to predict the number of awards won by students.

   - Information is provided on which of the three school programs the student takes part in (*General*, *Vocational* or *Academic*).
   - Also we are given the mathematics test score.

   State the mathematical formula used to predict the numer of awards won.
   You can denote ***progAcademic***, ***progVocational*** *and* ***math*** *as* x_1,x_2 *and* x_3 *respectively.*

```
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -5.2471     0.6585   -7.97  1.6e-15 ***
progAcademic     1.0839     0.3583    3.03   0.0025 **
progVocational   0.3698     0.4411    0.84   0.4018
math             0.0702     0.0106    6.62  3.6e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Comment on the significiance of each predictor variable used in the model?

4. (i) Use the model in Question (2) to predict the number of awards won by a general program student, with a maths score of 60.
   (ii) Use the model in Question (2) to predict the number of awards won by a vocational program student, with a maths score of 45.
   (iii) Use the model in Question (2) to predict the number of awards won by a academic program student, with a maths score of 80.

5. The R Code output given below is used to predict the number of Satelittes in the vicinity of a female crab.

   - We are given the width of each crab (in centimetres).
   - We are given information of the colour of the crab (Dark/ Not Dark)
   - We are given an assessment on the quality of the spine (GoodSpine/ Not GoodSpine)

State the mathematical formula used to predict the number of satelittes.
You can denote ***Width***, ***Darkyes*** *and* ***GoodSpineyes*** *as* x_1,x_2 *and* x_3 *respectively.*
Comment on the significance of each variable.

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.820088   0.570859  -4.940 7.81e-07 ***
Width         0.149196   0.020753   7.189 6.52e-13 ***
Darkyes      -0.265665   0.104972  -2.531   0.0114 *
GoodSpineyes -0.002041   0.097990  -0.021   0.9834
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

6. This question is an extension of Question (5), using only ***Width*** and ***Darkyes*** as predictor variables.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.82022    0.57073  -4.941 7.76e-07 ***
Width        0.14917    0.02072   7.201 5.98e-13 ***
Darkyes     -0.26518    0.10235  -2.591  0.00957 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

   (i) Use the model to predict the number of satellites in the vicinity of a lightskinned crab with a width of 20cms

   (ii) Use the model to predict the number of satellites in the vicinity of a lightskinned crab with a width of 30cms

   (iii) Use the model to predict the number of satellites in the vicinity of a dark-skinned crab with a width of 20cms

7. What is Zero Inflation? Explain the Modeling Process for a Zero Inflated Model. Give an Example of Zero-Inflated Count Process. *Support your answer with a sketch, if necessary.*

8. Describe a situation whereby Negative Binomial Regression Models would be used instead of Poisson Models.

9. What is Zero Inflation? Give an example of a Zero Inflated Count Process

10. What is Zero Truncation? Give an example of a Zero Truncated Count Process

11. in the context of modelling count variables, what is Vuong Test for?