Trends of the number of infected cases in female age groups

Abstract: The simulation and analysis of the final project is based on the data of cases infected with Coronavirus in Florida State. Among the 11 age groups, the cases of female infected from age groups 6 to 11 are studied here in order to find the relationship between the number of infected cases and time (each day) for each age group. Nonlinear methods are applied to fit the trends of daily number and total number of infected cases. The simulation results show: (1) for the trends of daily number of infected cases versus day, both methods used in this project fit the model well while the modified quadratic Michaellis-Menten (MM) method slightly outperforms another method indicated by MSE; (2) for the trends of total number of infected cases versus day, the quadratic MM model works better than Poisson model.

1. Preliminary Analysis

After data cleaning, the scatterplots are used to demonstrate the potential mathematical relations between the daily number or the total number of infected cases and time (day) within each age group, shown as below in Fig.1 and Fig.2.

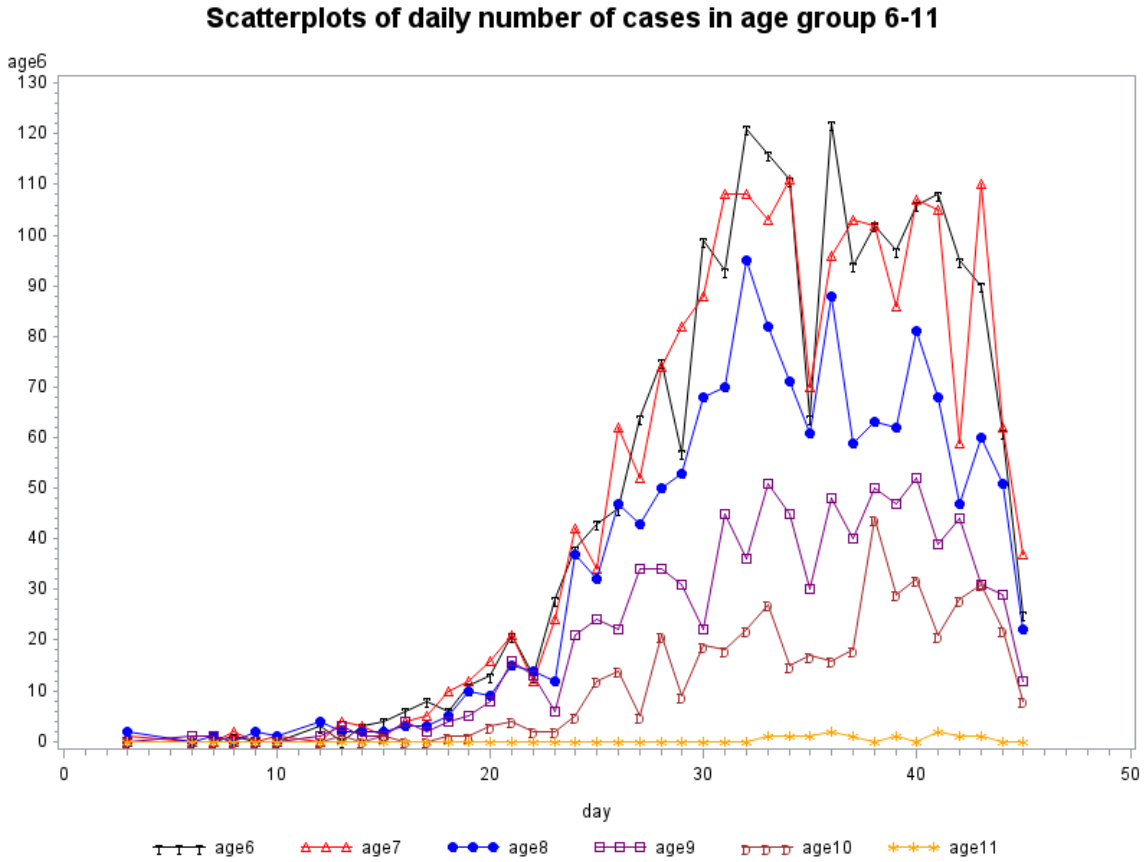

Fig. 1 Scatterplot of the daily number of cases Versus day within each age group

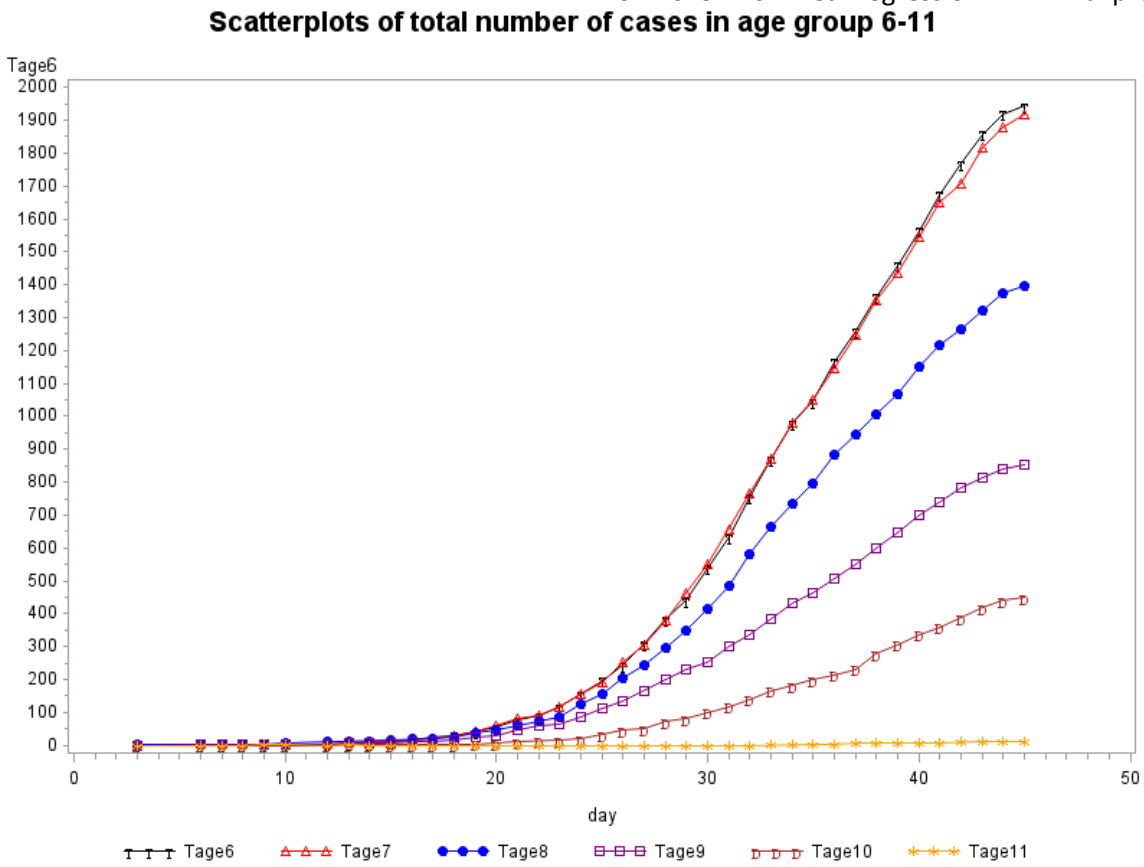

Fig. 2 Scatterplot of the total number of cases Versus day within each age group

From Fig.1, two potential models seem to be appropriate to fit the daily number of cases versus day, represented as below, respectively.
Model 1: the number of cases = 1 / (theta1 + theta2*day + theta3*day^2),
Model 2: the number of cases = theta1*day / (theta2 + day + theta3*day^2).
Note that model 2 is the quadratic Michaellis-Menten model. Both models have three parameters (theta1, theta2, theta3) to initialize.
From Fig.2, Poisson and the quadratic MM models can be appropriate to fit the total number of cases versus day.

2. Initialization of parameters

According to the data given, the initial parameters for model 1 and 2 are found (1) theta1 = -18.36, theta2=-18.69, theta3=-0.0145 for model 1; (2) theta1= 1, theta2 = -0.00546, theta3=-0.000238 for model 2, when using proc NLIN. However, Poisson model doesn't need the initialization of parameters.

3. Model fitting

For the daily number of cases versus day, through proc NLIN, the results of both models related to the daily number of cases are compared. Since in age group 11 there are only 11 observations, it's insufficient to find the potential trend behind the data. Through comparing the results, we can observe that: (1) both models are significant indicated by the F-test for each age group; (2) both models mostly violate the assumptions of constant variance and normality; (3) model 2 (MM model) seems to outperform model 1 by comparing their MSEs. Hence, the model 2 is supposed to be the final model for the daily number of cases versus day.

For the total number of cases versus day, Poisson and the quadratic MM models are introduced to fit the potential trends for each age group. The Poisson model for each age group perform well without

overdispersion or underdispersion indicated by scale=1, however, from the trend plots of the predicted total number of cases versus day in Fig.4 and Fig.5 , the MM model seems to be better than Poisson model.

4. Trend plots

For the daily number of cases versus day, based on the outputs of MM model (model 2), the potential trend of the daily number of cases versus day within each age group is indicated as below in Fig.3. Note that in Fig.3, age6, age7, age8, age9, age10, age11 denote the observed daily number of cases within each age group; p6, p7, p8, p9, p10 represent the corresponding predicted daily number of cases within each age group. For age group 11, due to the lack of data, only observed daily number of cases is shown, and there is no trends found for this age group.

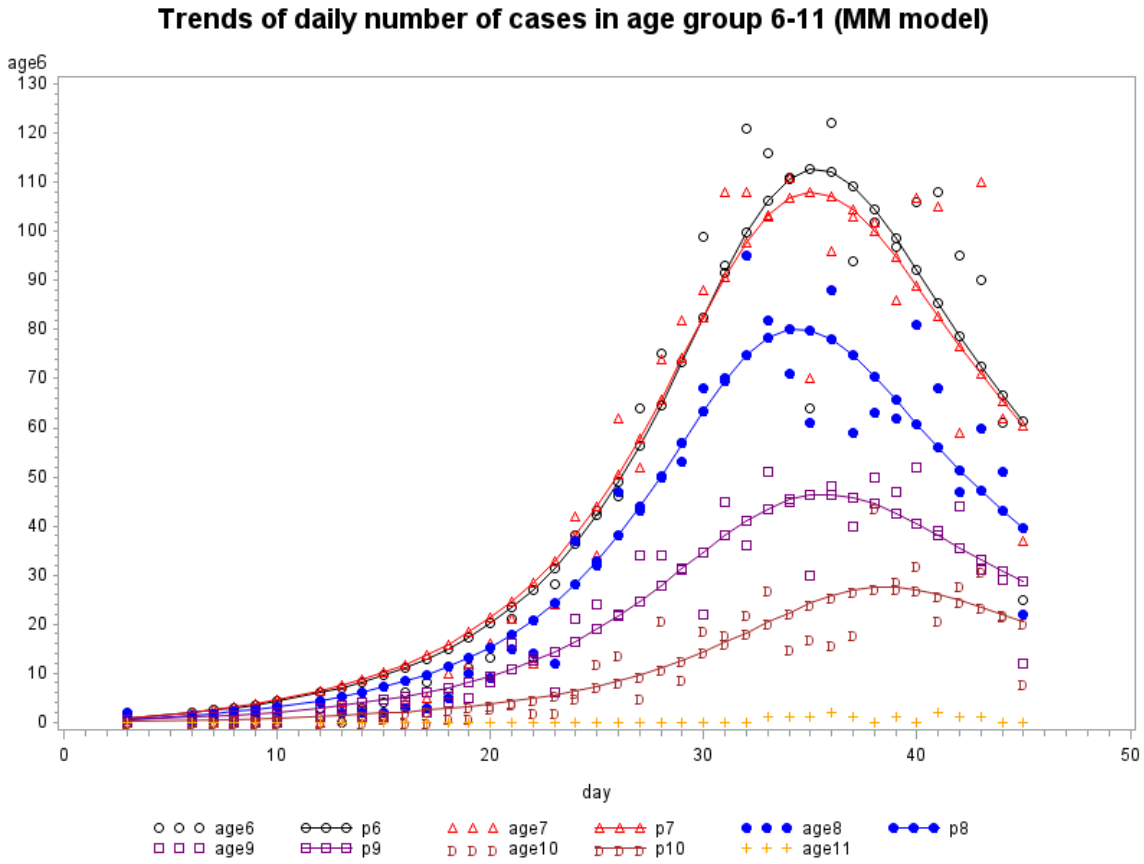

Fig.3 Trend of daily number of infected cases versus day within each age group

For the total number of cases versus day, both trend plots of Poisson and the MM models are indicated as below in Fig.4 and Fig.5. Note that Tage 6-11 denote the observed total number of cases versus day within each age group; predicted values and p6-11 represent the corresponding predicted total number of cases within each age group in Poisson and MM models.

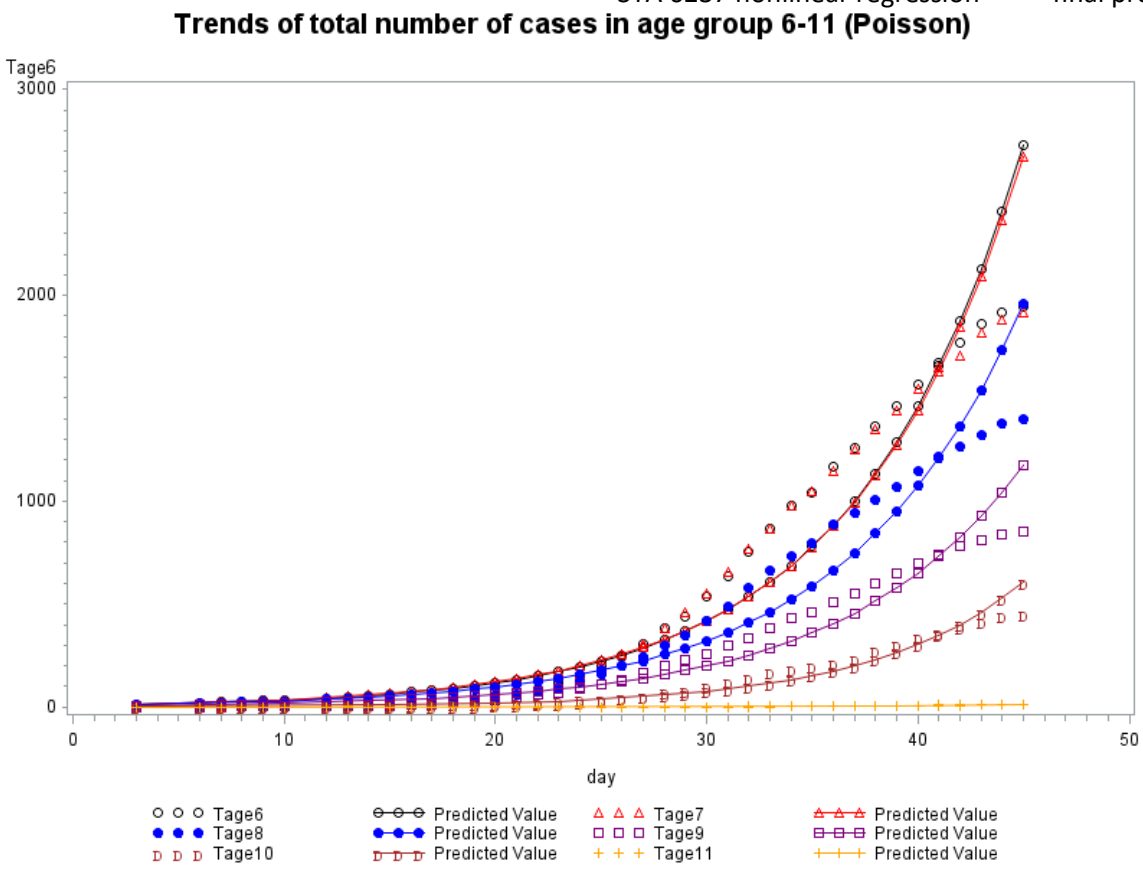

Fig.4 Trend of daily number of infected cases versus day within each age group

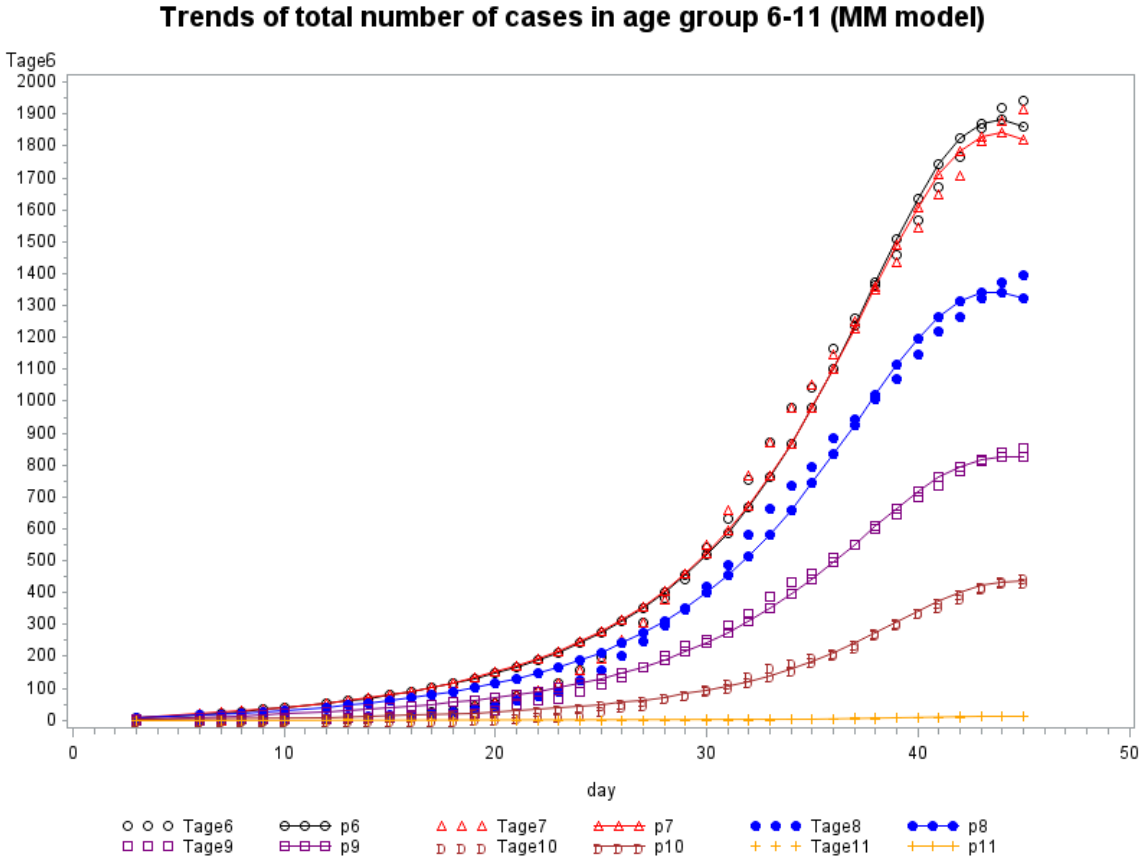

Fig.5 Trend of total number of infected cases versus day within each age group