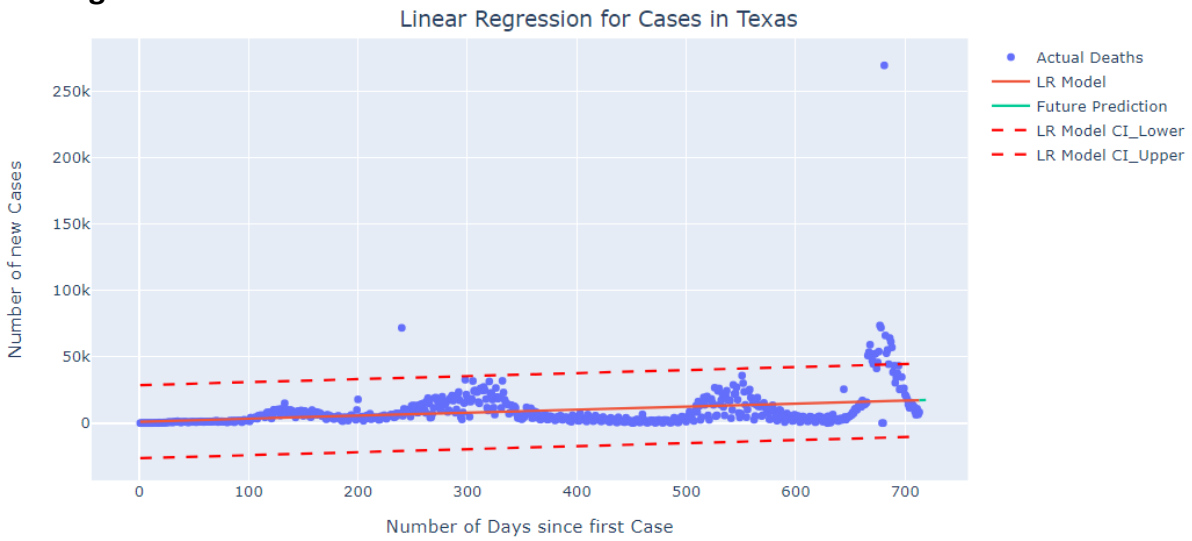
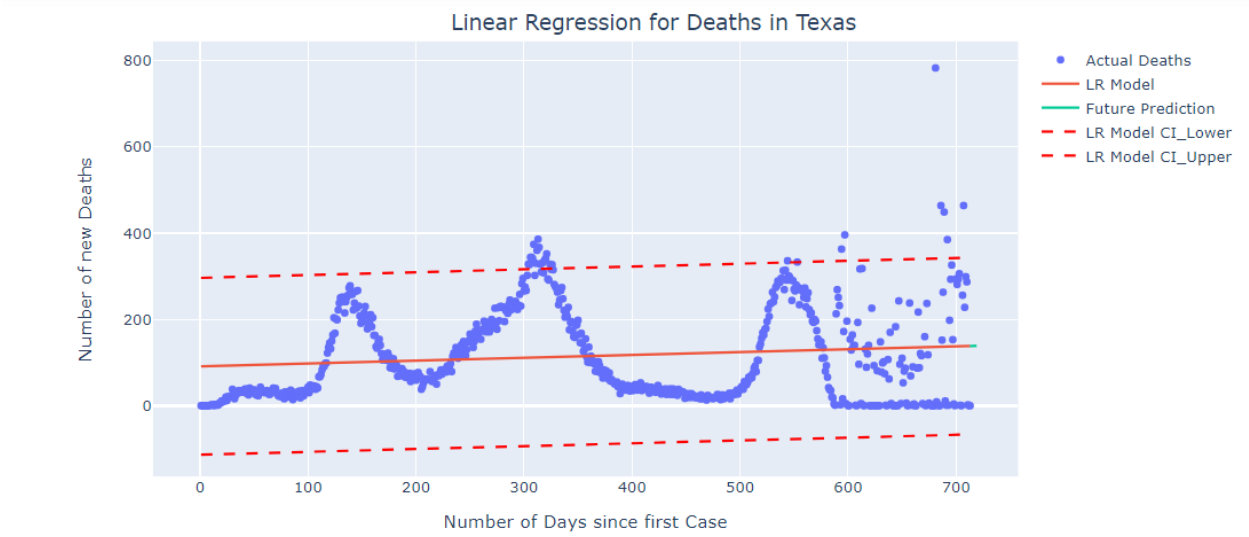
Utilize Linear and Non-Linear regression models to compare trends for a Texas state.

Linear Regression for Cases – Texas State



RMSE: 14016.649068165274

Linear Regression for Deaths – Texas State

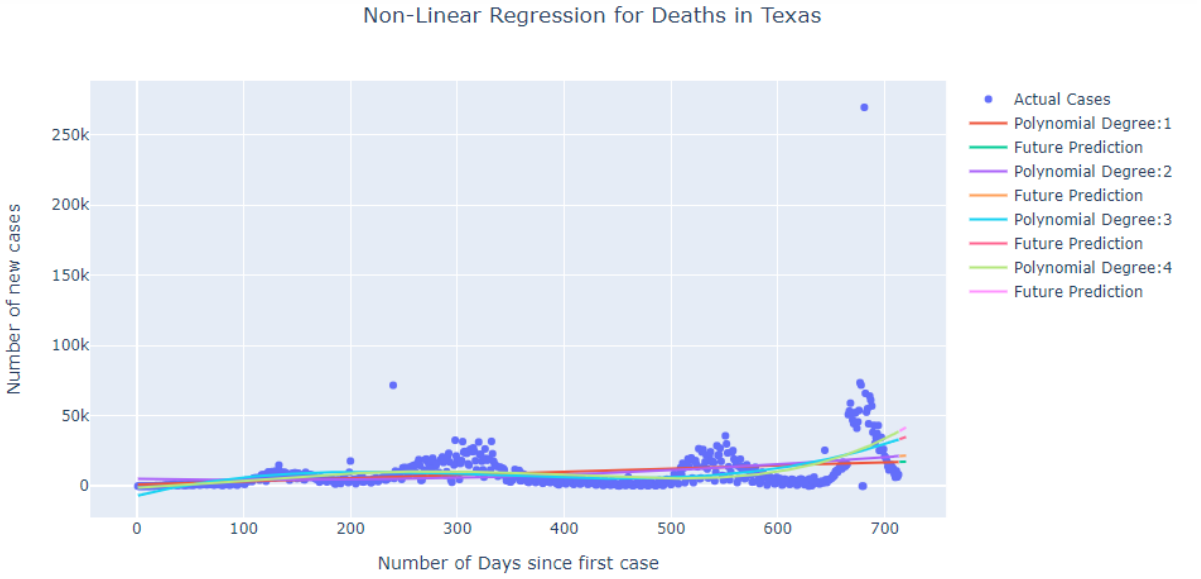


RMSE: 104.41837115684291

Non-Linear Regression for Cases – Texas State

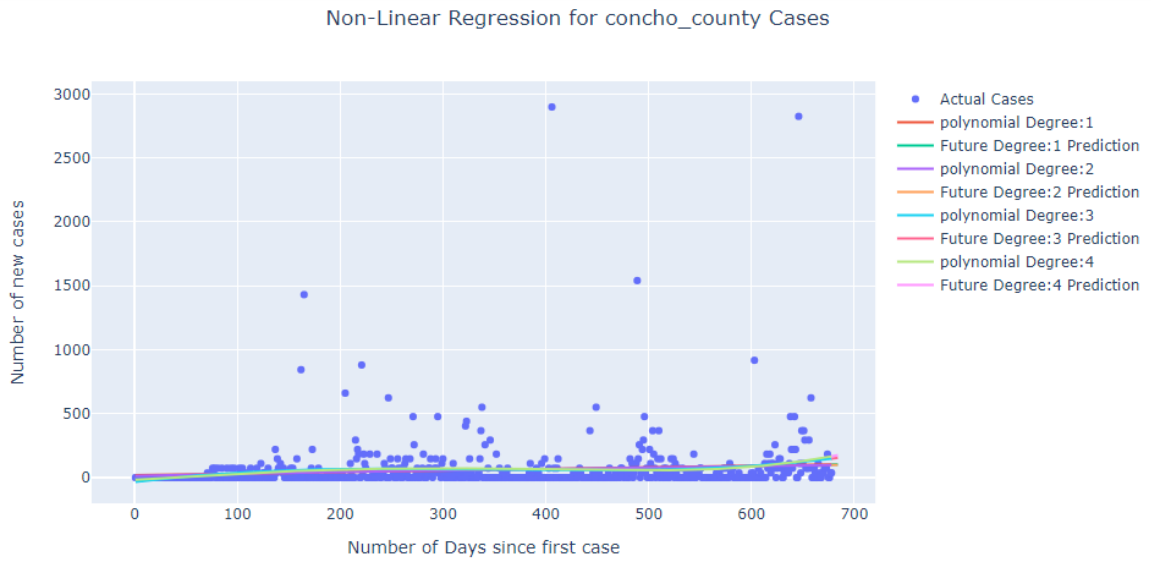
RMSE for degree = 1 is 14016.649068165274
RMSE for degree = 2 is 13897.354934418747
RMSE for degree = 3 is 13143.84481433265
RMSE for degree = 4 is 13005.77702809882

Non-Linear Regression for Deaths – Texas State

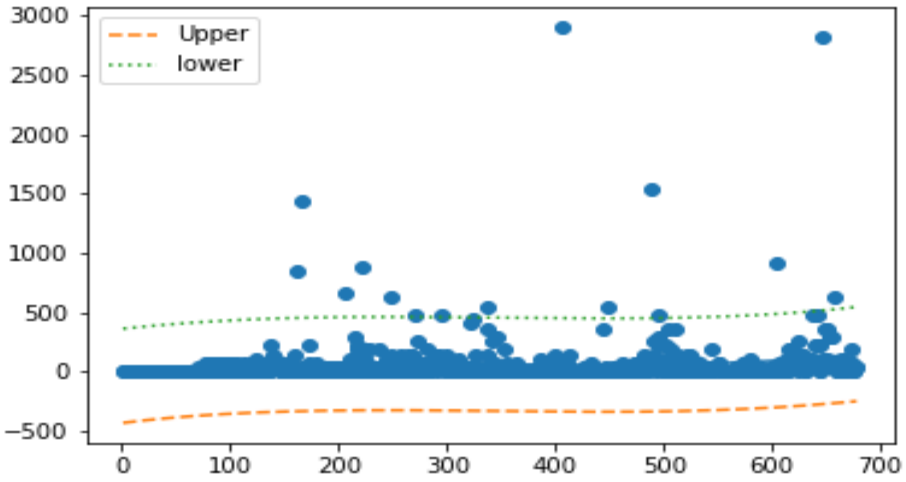


Identify which counties are most at risk. Model for top 5 counties with cases within a state

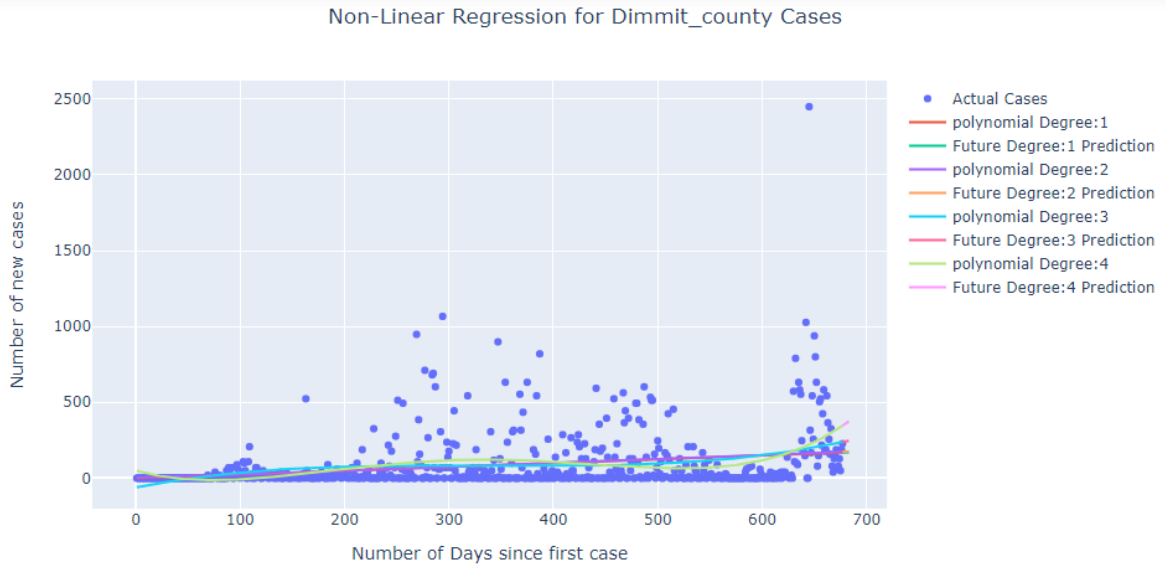
Non-Linear Regression for Concho County Cases



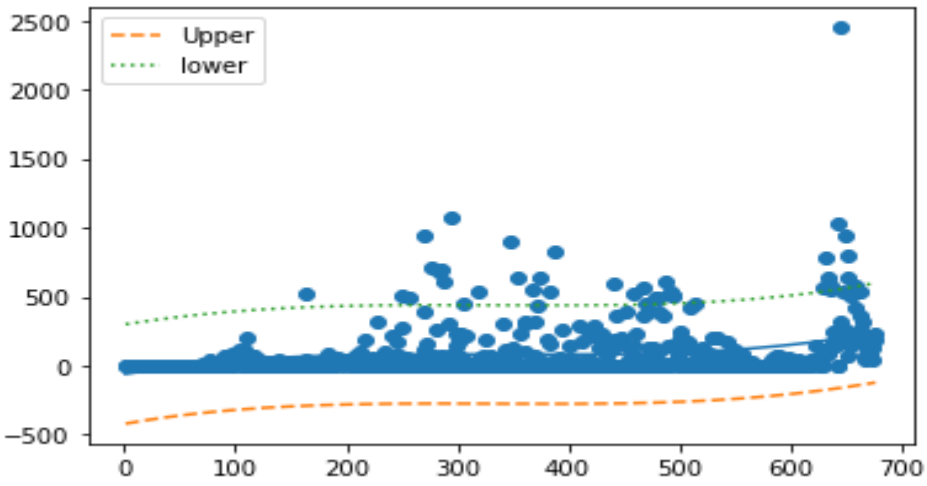
Confidence Interval



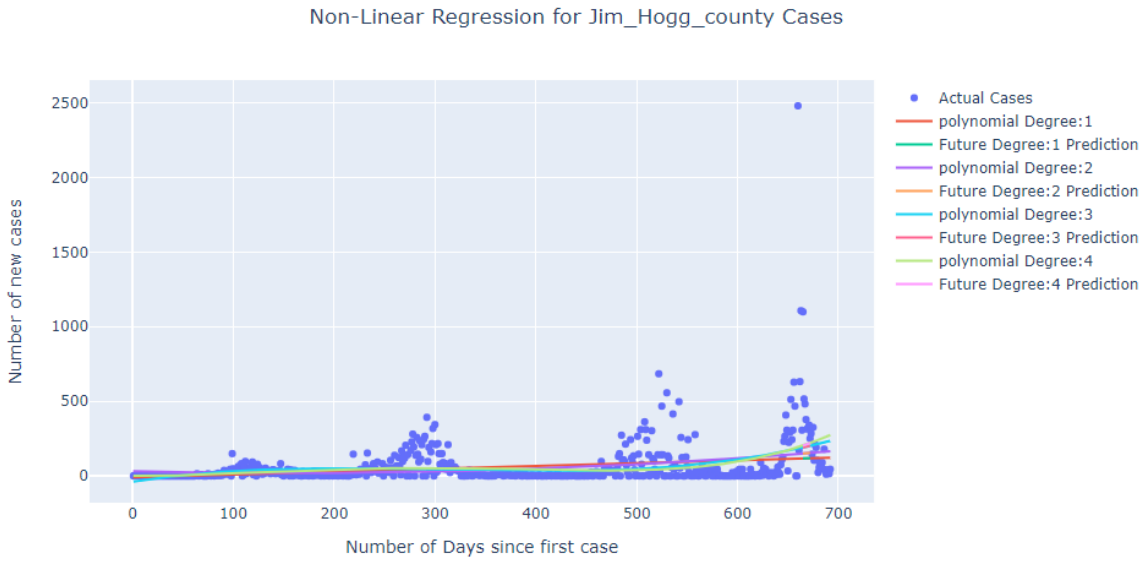
Non-Linear Regression for Dimmit County Cases



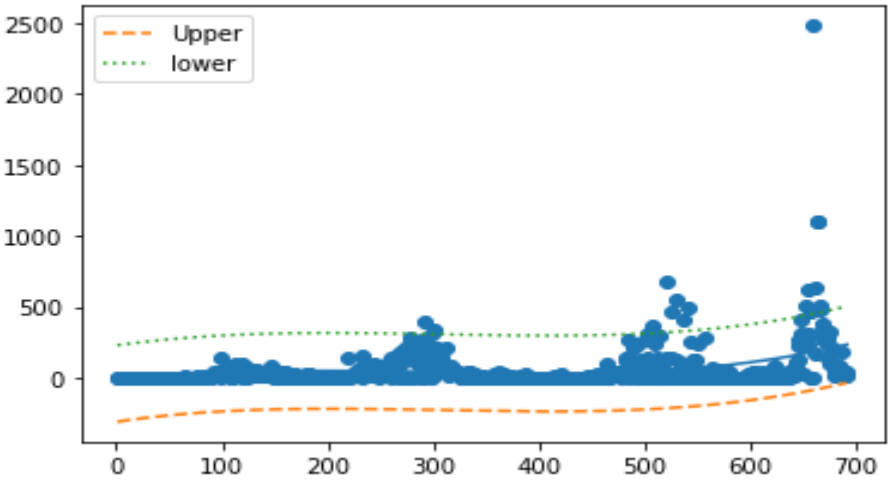
Confidence Interval



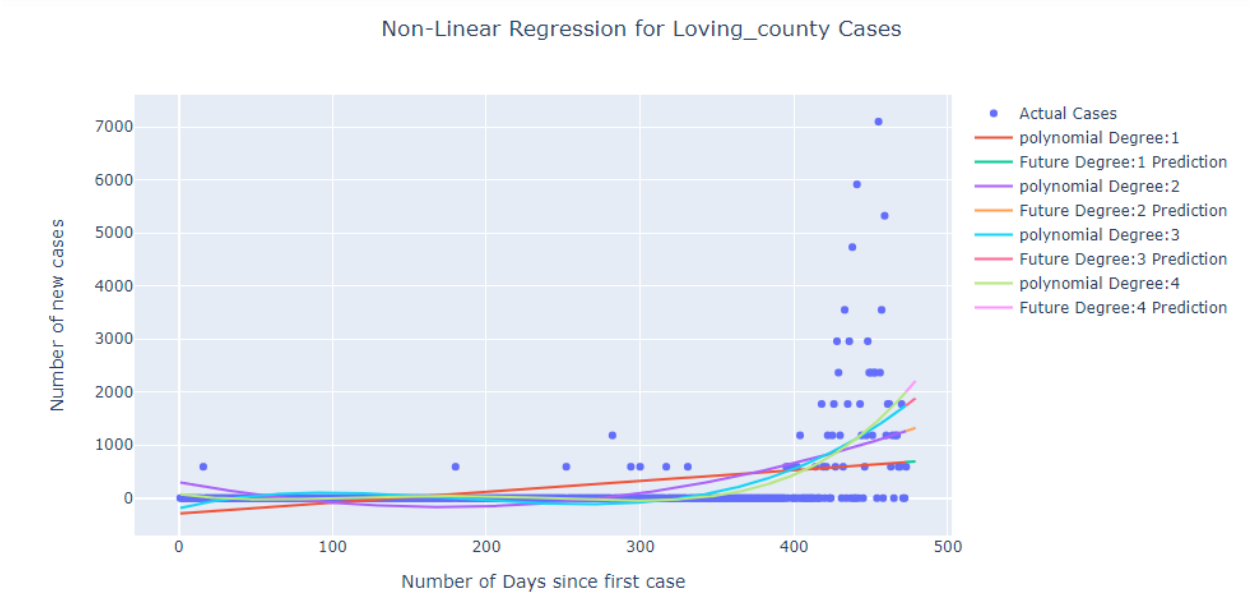
Non-Linear Regression for Jim Hogg County Cases



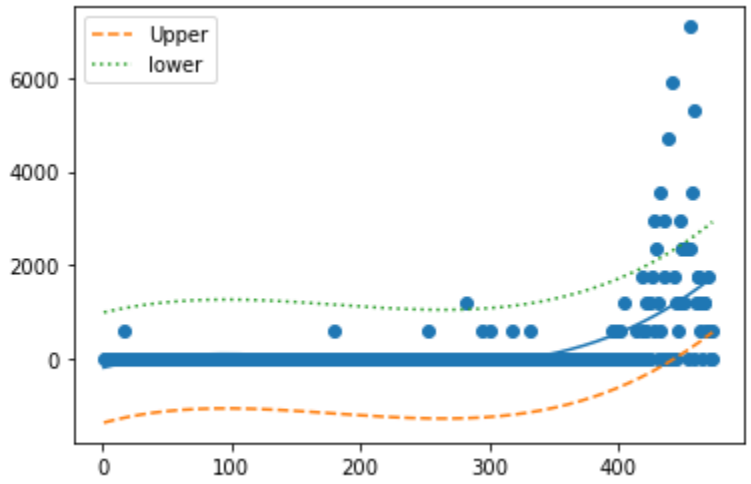
Confidence Interval



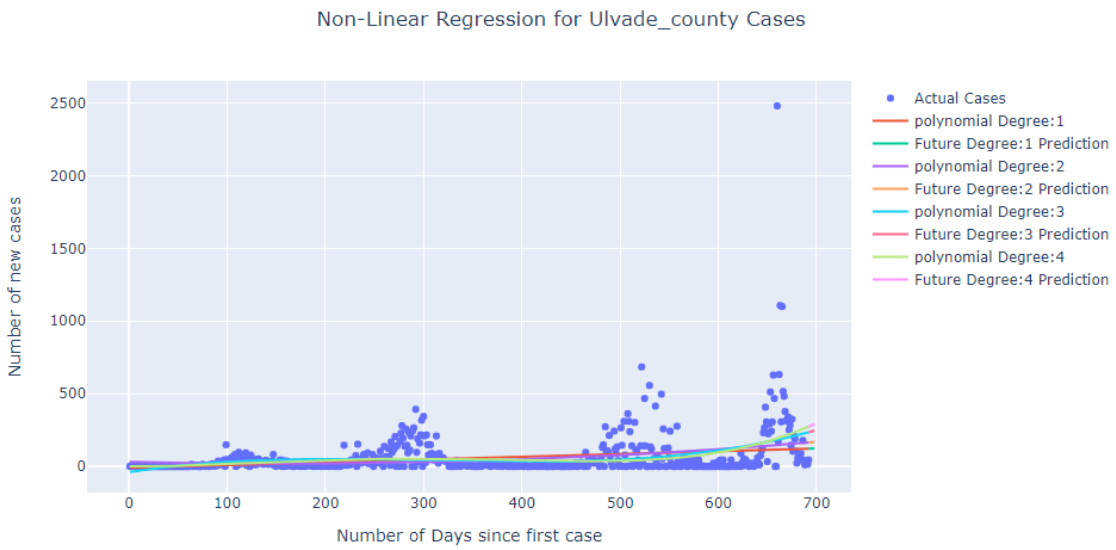
Non-Linear Regression for Loving County Cases



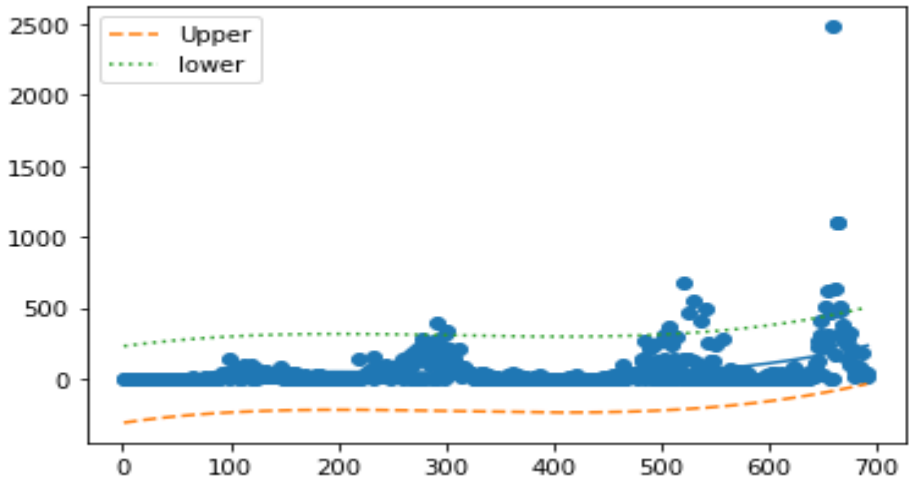
Confidence Interval



Non-Linear Regression for Ulvade County Cases



Confidence Interval



Observations

By looking at the trend lines and prediction paths, we can say that Dimmit County and Concho County are at high risk as the cases count is predicted to increase on a high scale when compared to other counties

Perform hypothesis tests on questions identified in Stage II

Data considered for 3 States - Alabama, Alaska, and Arkansas.

Null Hypothesis : Death rate of seniors is more than other age groups

Alternative Hypothesis : Death rate of seniors is ***not*** more than other age groups

In [56]:
```
stats.ttest_ind(a=data_group_st_age['Age_85over'], b= data_group_st_age['Death'],equal_var=False)
```

Out[56]:
```
Ttest_indResult(statistic=1.3143882000665275, pvalue=0.31735670321180154)
```

The pvalue in this case is greater than significance level (0.05). Hence, we cannot reject the null hypothesis.

Null Hypothesis : Death rate of children is more than other age groups

Alternative Hypothesis : Death rate of children is ***not*** more than other age groups

In [57]:
```
stats.ttest_ind(a=data_group_st_age['Age_under5'], b= data_group_st_age['Death'],equal_var=False)
```

Out[57]:
```
Ttest_indResult(statistic=1.6178485047620426, pvalue=0.24690373079360323)
```

The pvalue in this case is greater than significance level (0.05). Hence, we cannot reject the null hypothesis.

Null Hypothesis : Males are more prone to covid

Alternative Hypothesis : Males are ***not*** more prone to covid

In [58]:
```
stats.ttest_ind(a=data_group_st_gender['Total_Male'], b= data_group_st_age['Cases'],equal_var=False)
```

Out[58]:
```
Ttest_indResult(statistic=1.3003478790962522, pvalue=0.3164960946318385)
```

The pvalue in this case is greater than significance level (0.05). Hence, we cannot reject the null hypothesis.

Null Hypothesis : Females are more prone to covid

Alternative Hypothesis : Females are ***not*** more prone to covid

In [59]:
```
stats.ttest_ind(a=data_group_st_gender['Total_Female'], b= data_group_st_age['Cases'],equal_var=False)
```

Out[59]:
```
Ttest_indResult(statistic=1.2669916622097117, pvalue=0.3271177552979792)
```

The pvalue in this case is greater than significance level (0.05). Hence, we cannot reject the null hypothesis.