# Project 3 in Linear and Logistic Regression

El-Tayeb Bayomi, Nora Lüpkes

Lund University

May 27, 2020

# Introduction

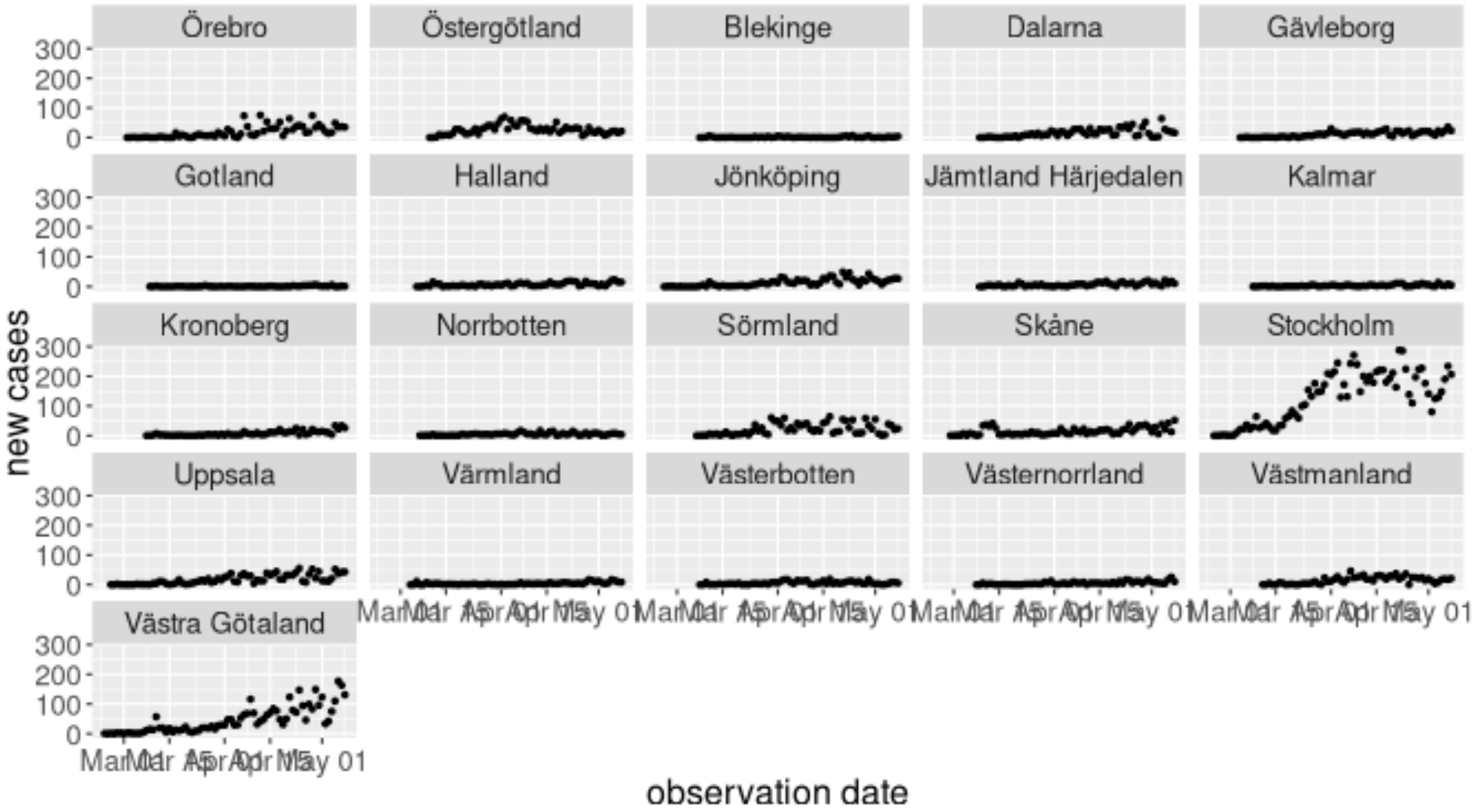
## Purpose and expectations of this project

- COVID-19
- Implementing what we have learnt throughout the course.
- Train presenting your independent work.

# Data structure

- obs_date, **The date (in R date format).**
- day_nbr, **Consecutive day number.**
- day_nbr_region, **Consecutive day number within region.**
- region **The name of the region.**
- population, **The population size.**
- new_cases, **the number of new confirmed cases.**

# Raw data


Örebro
Östergötland
Blekinge
Dalarna
Gävleborg
Gotland
Halland
Jönköping
Jämtland Härjedalen
Kalmar
Kronoberg
Norrbotten
Sörmland
Skåne
Stockholm
Uppsala
Värmland
Västerbotten
Västernorrland
Västmanland
Västra Götaland
300
200
100
0
new cases
May 01
observation date

# NB/PO-models

**Model**:

`new_cases` $\sim$ `(day_nbr_region) * (day_nbr) * obs_date + region`

`population` as offset variable

Family = negative binomial or poisson ?

# step()

**Model**:

`new_cases` $\sim$ `(day_nbr_region) * (day_nbr) * obs_date + region`

`population` as offset variable
Family = negative binomial

| | Df | BIC |
|---|---|---|
| none | | 110323 |
| - day_nbr_region:day_nbr:obs_date | 1 | 110567 |
| - region | 19 | 131846 |

# step()

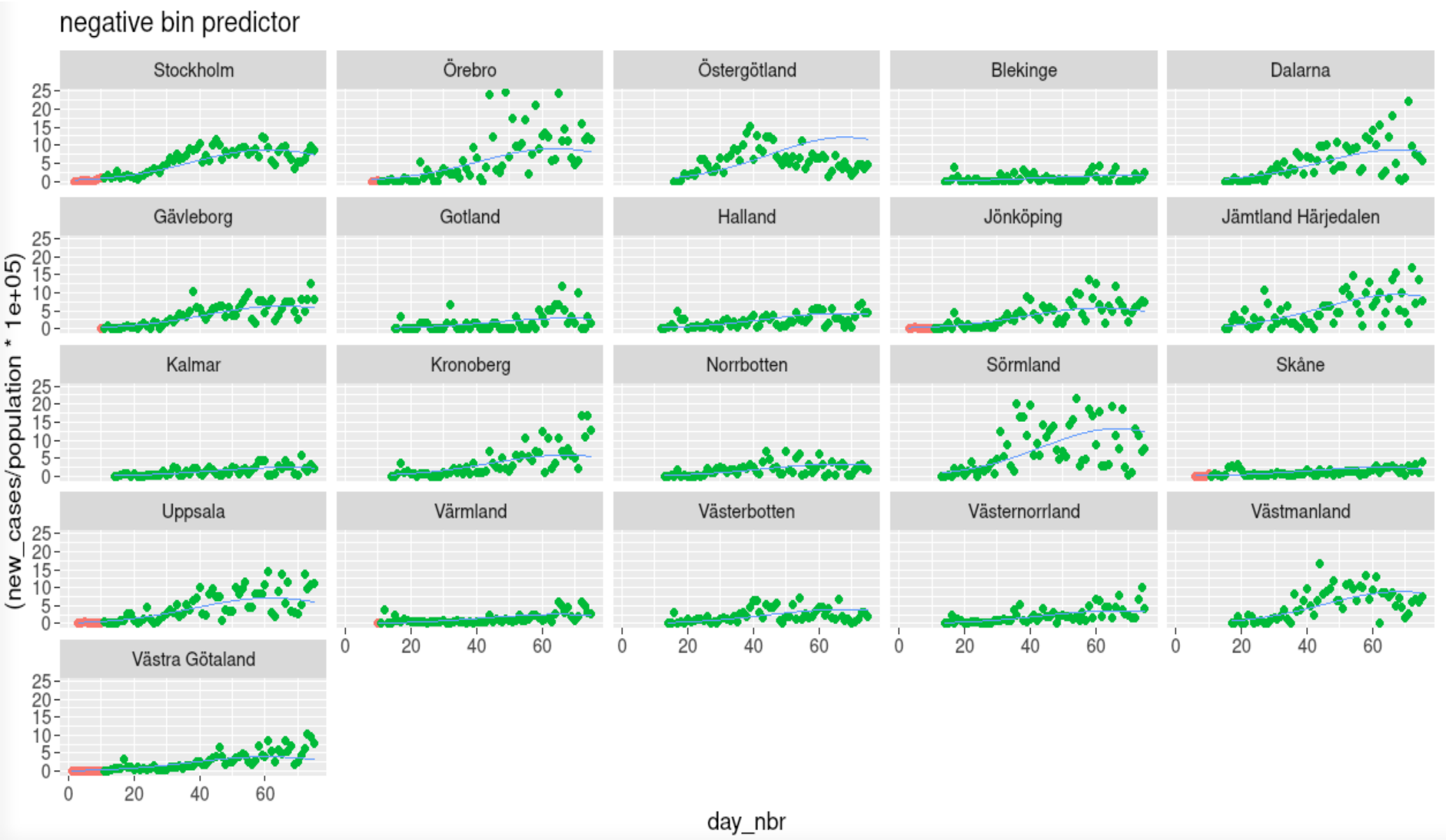
**Model**:

`new_cases` $\sim$ `(day_nbr_region) * (day_nbr) * obs_date + region`

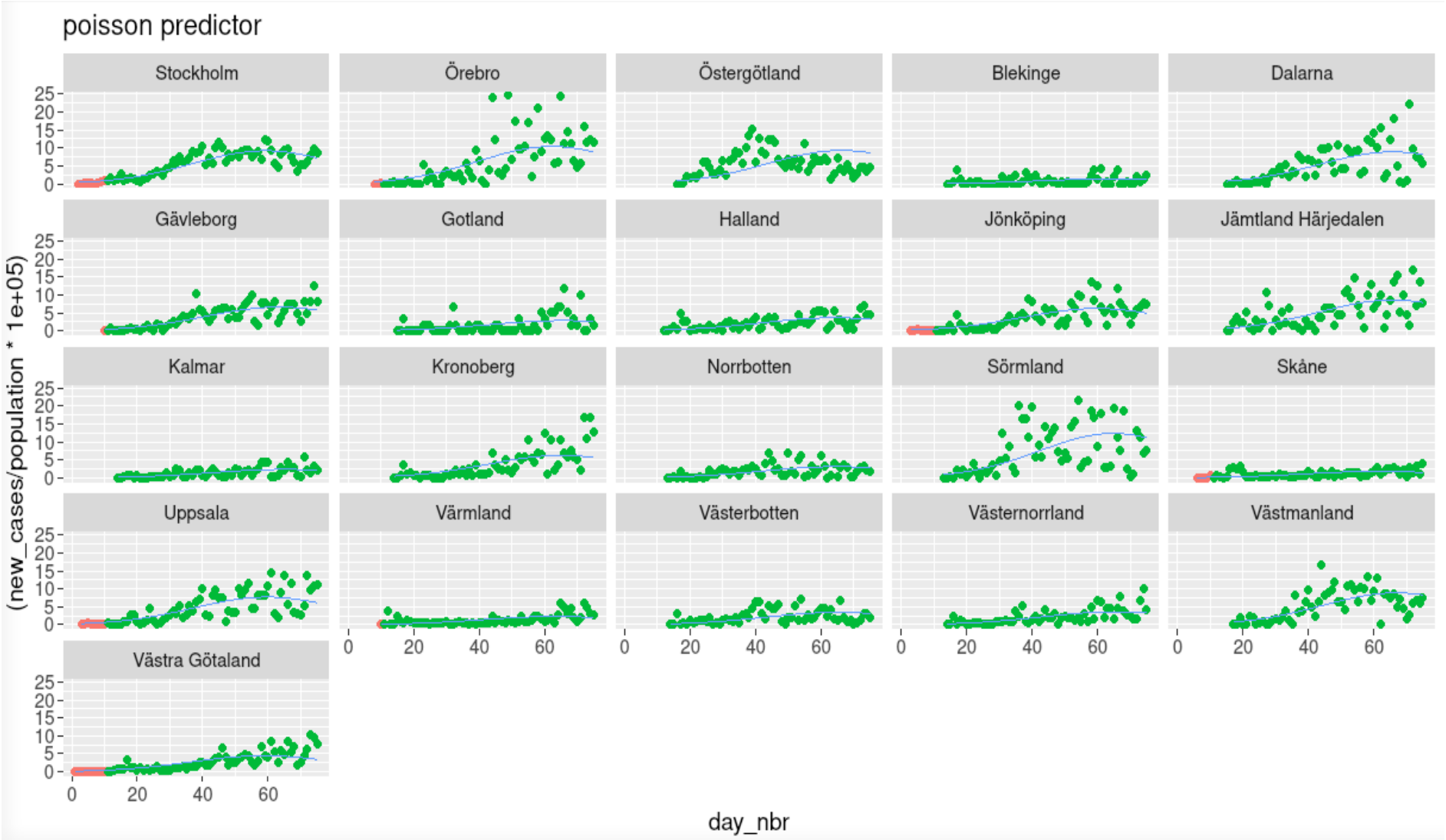
`population` as offset variable
Family = negative binomial

| | Df | BIC |
|---|---|---|
| none | | **110323** |
| - day_nbr_region:day_nbr:obs_ date | 1 | 110567 |
| - region | 19 | 131846 |

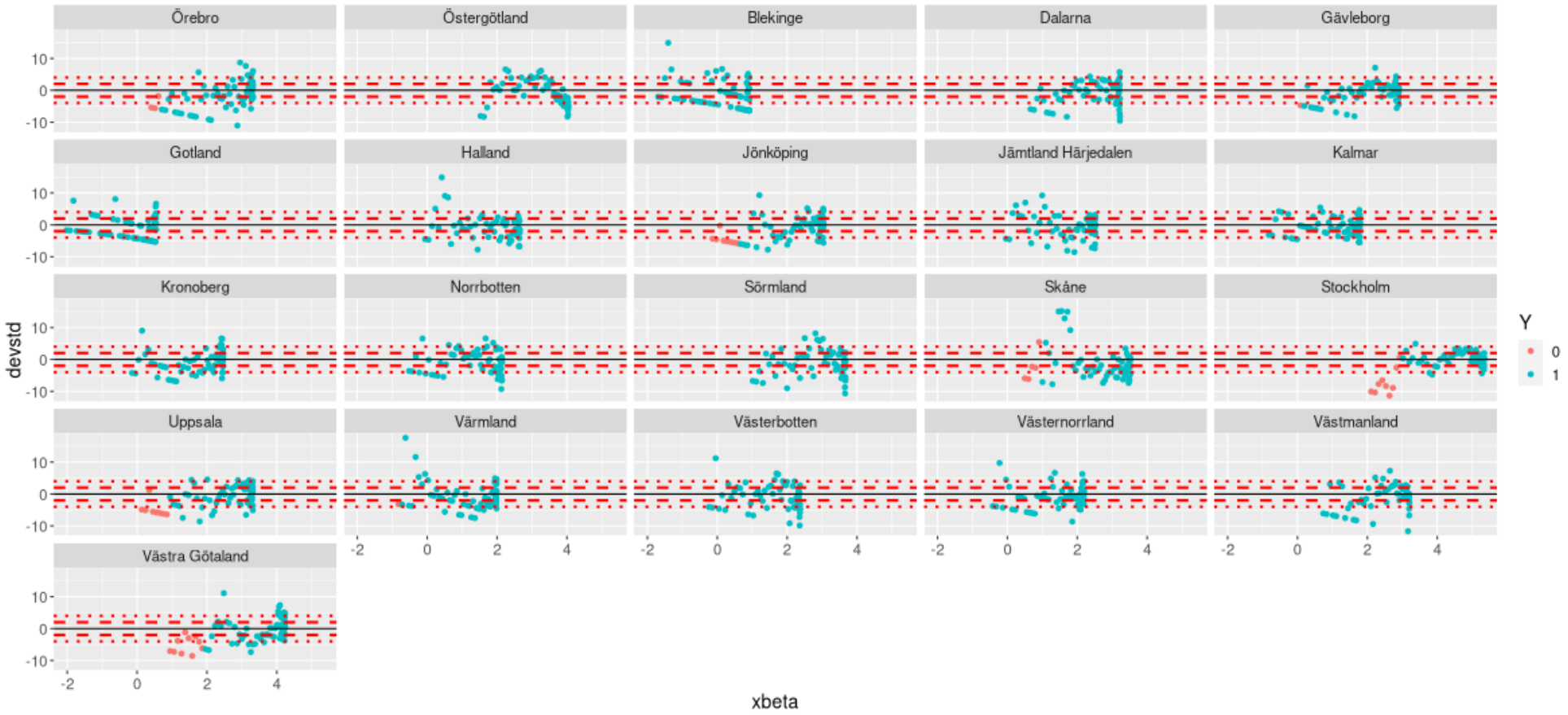
# Negative binomial predictions



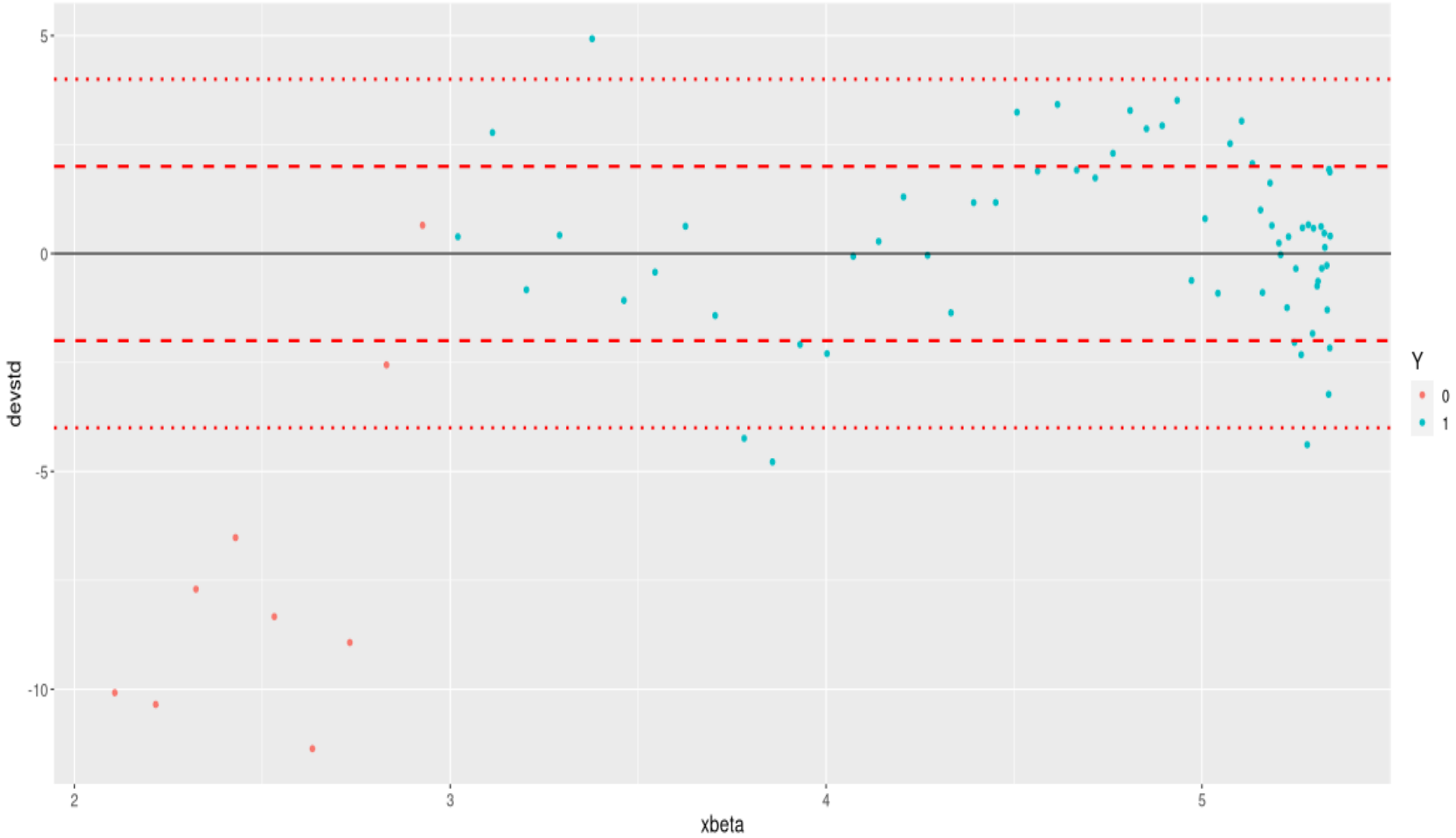
# Poisson predictions



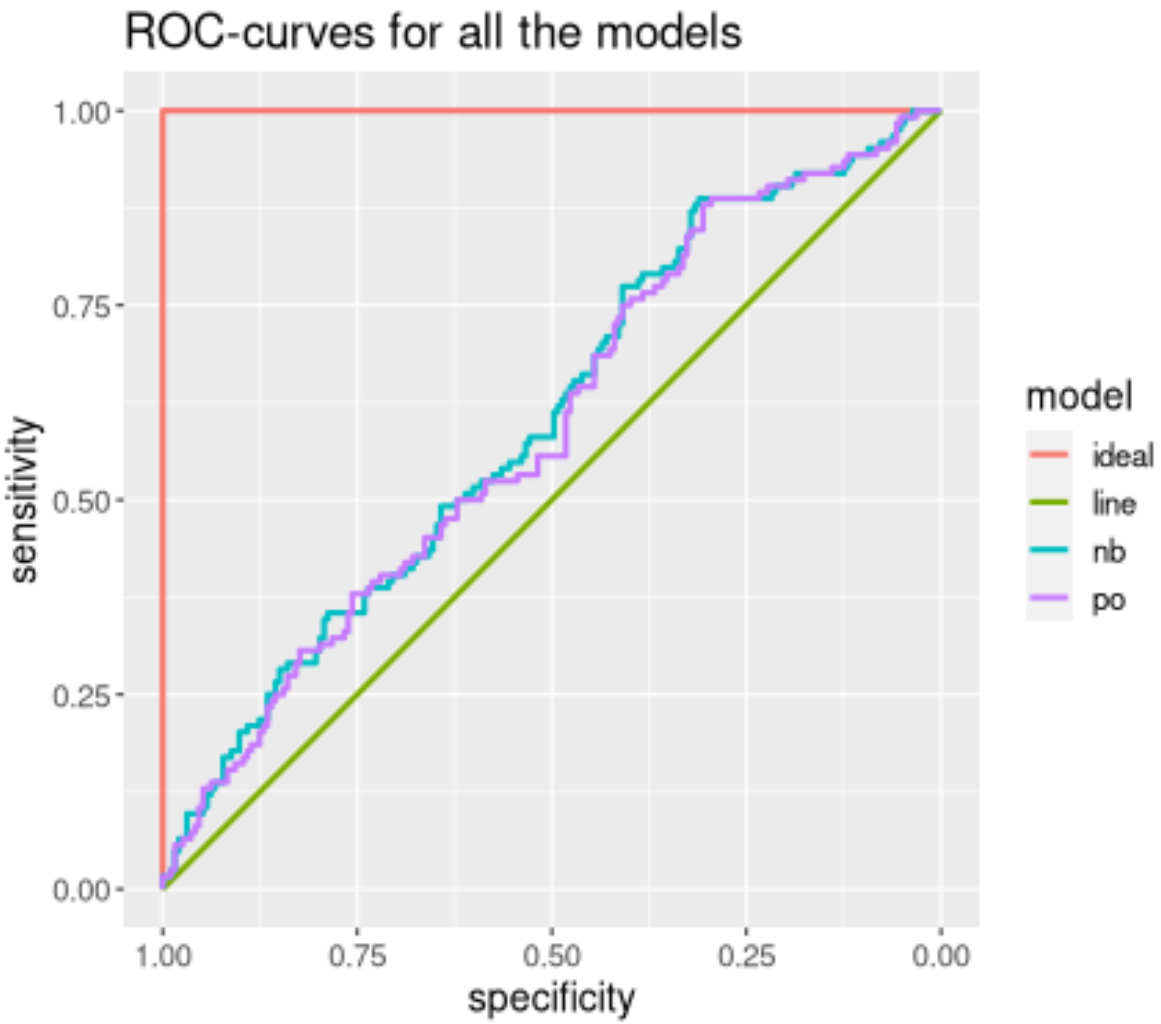
# Negative binomial - Deviance standarized residuals



# Deviance standarized residuals - Stockholm



# ROC

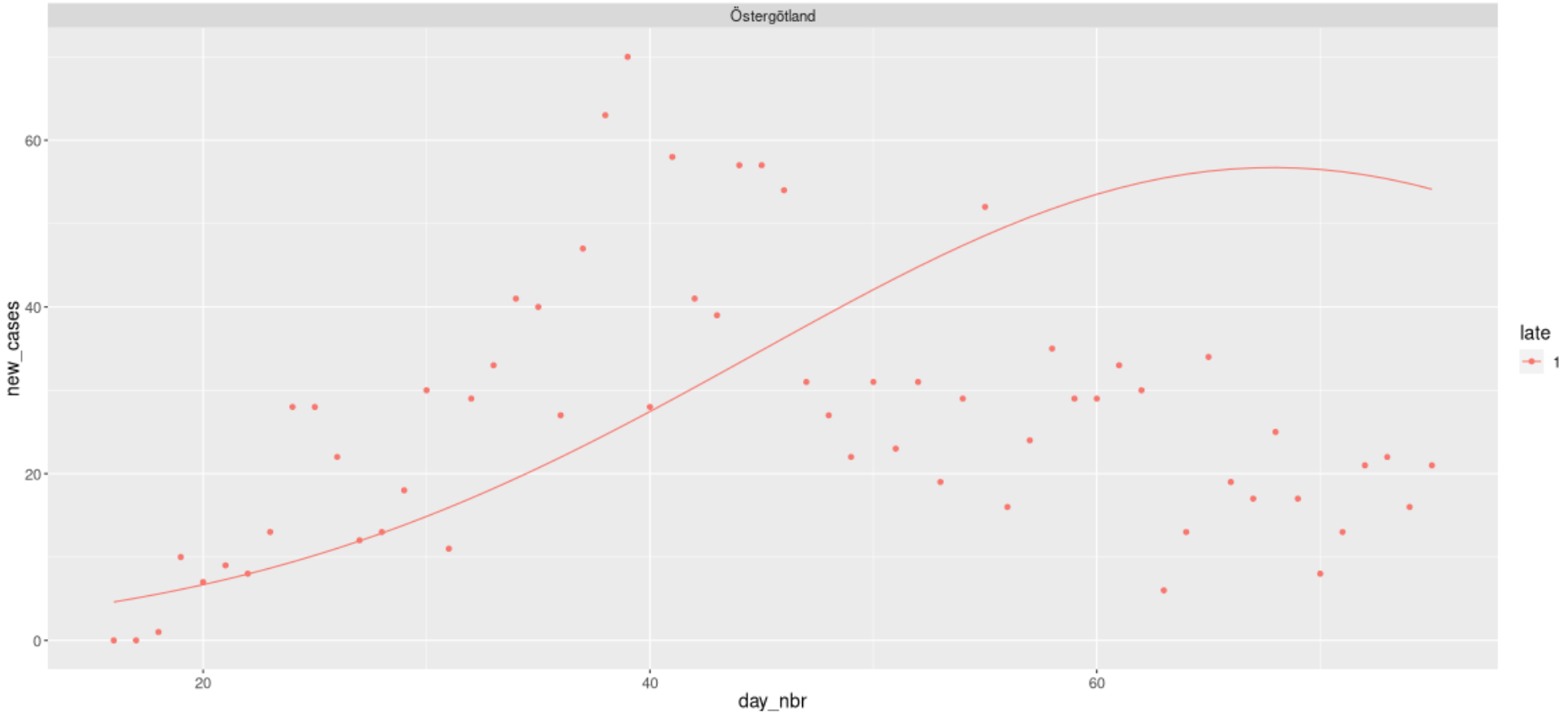

ROC-curves for all the models
sensitivity
1.00
0.75
0.50
0.25
0.00
1.00
0.75
0.50
0.25
0.00
specificity
model
ideal
line
nb
po

# AUC

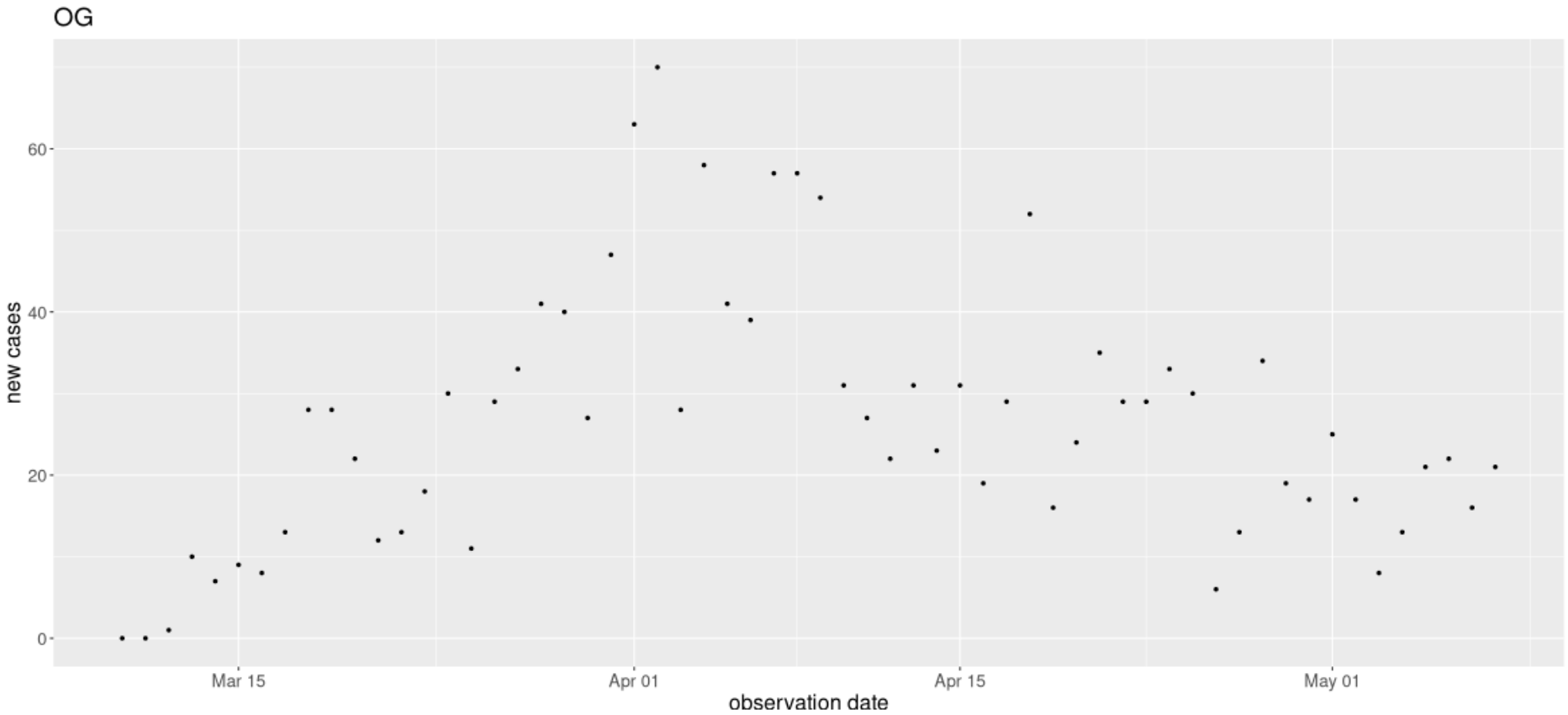
```
roc.test(mod_po.roc, mod_nb.roc)
```

| AUC of Poisson | AUC of negative binomial |
|---|---|
| 0.5932225 | 0.6014123 |

# Östergötland - continuing work?



# Östergötland - continuing work?

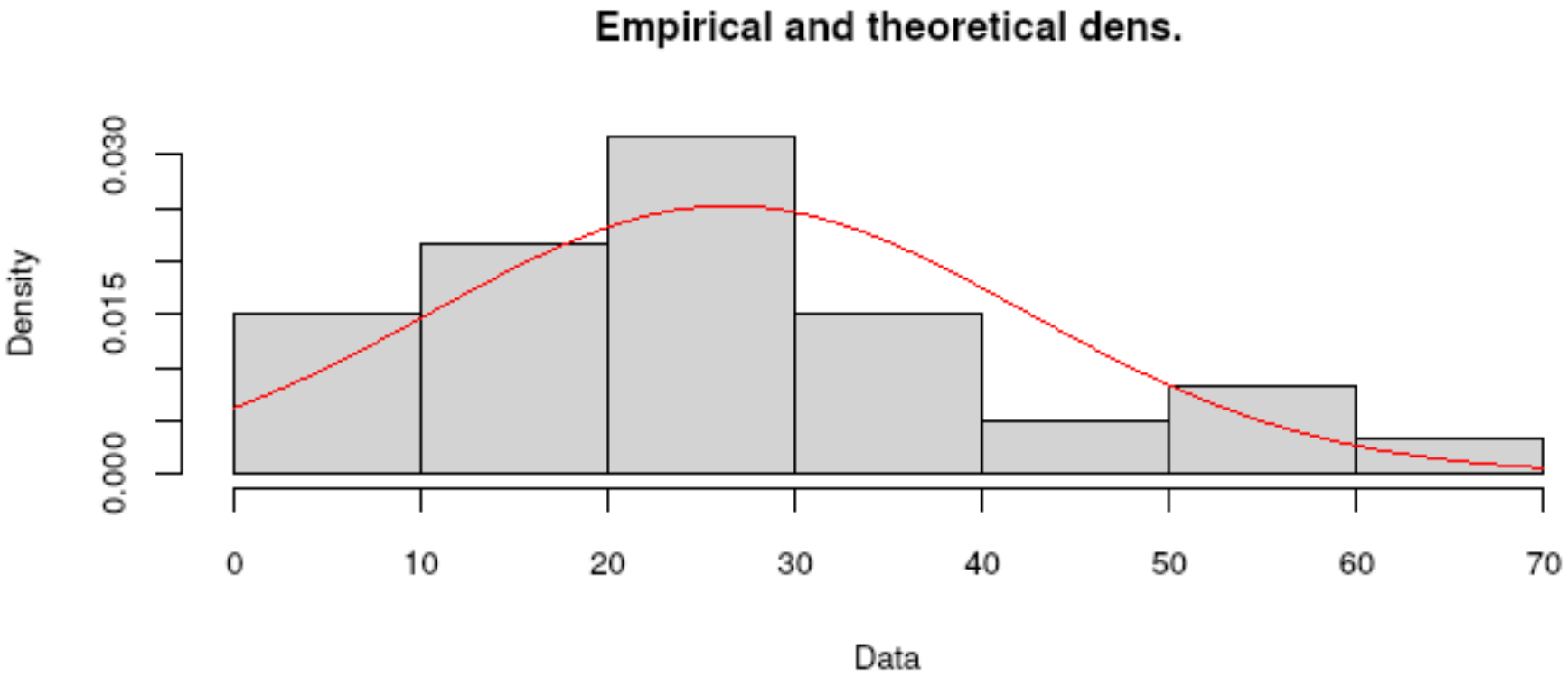


# Fit a Gaussian curve

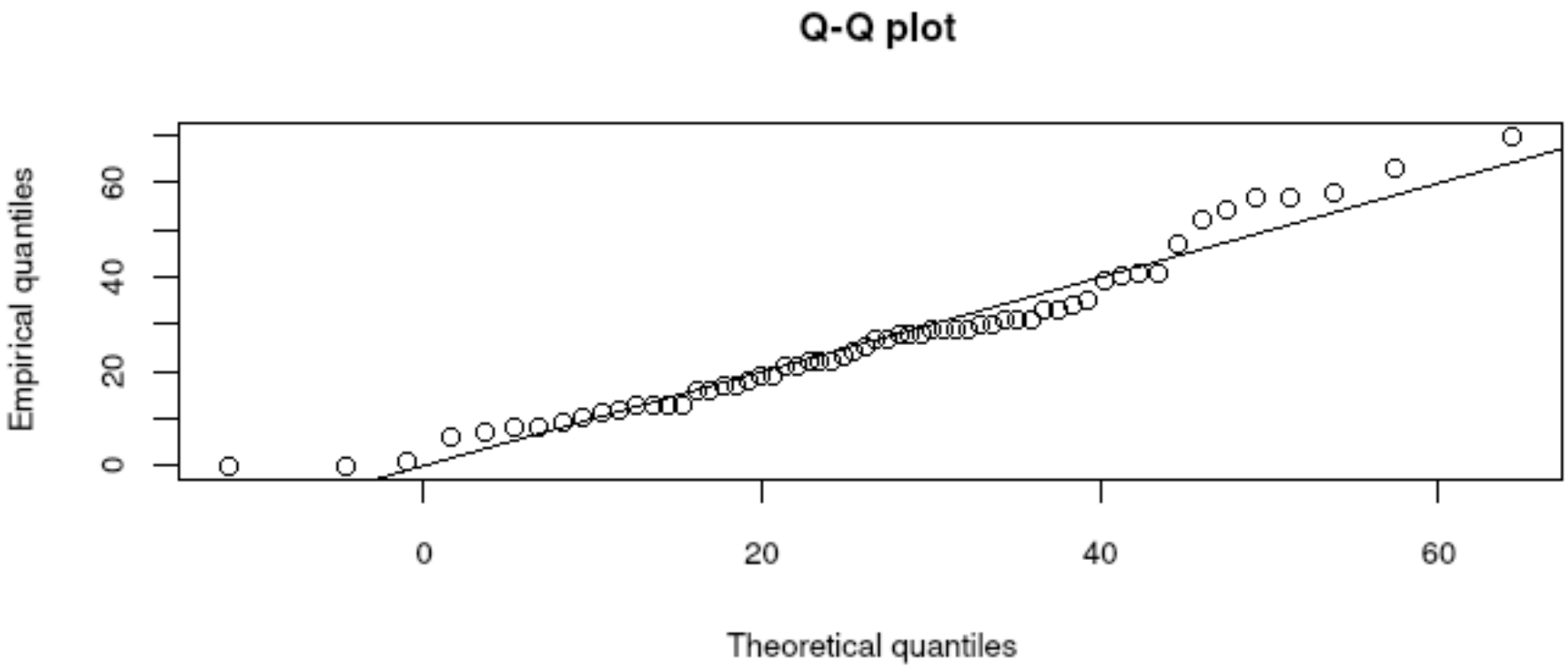
**Code**:

```
library(fitdistrplus)
FIT <- fitdist(fhm.data.og$new_cases, "norm")
class(FIT)
plot(FIT)
```

# Östergötland - continuing work?



# Östergötland - continuing work?



# Outlook

Peak present?

$\Rightarrow$ Choose model family

Thank you!