LM-GLM. Second Deliverable. 2017-18

Note:

In this deliverable, you have to do the exercise that appears in the section WORDING, and posteriorly, to answer the questions that appear in the section QUESTIONS. You have to answer the questions in ATENEA "2a entrega", you have an unlimited number of trials when answering.

WORDING

One hundred days where randomly selected in a given year. For each one of the days, the number of visits in an emergency department of an hospital (y) were recorded, jointly with an index of the level of pollen in the atmosphere that day (x) and the mean of that day temperature (t). The data appear in the file: "Ficheros de datos de la 2a entrega" in Atenea. (Take into account that they are simulated data and thus, you can not trust in the results).

The objective is to detect if there is a significant relationship between the number of visits and the pollen level, and the number of visits and the temperature. To that end, three Poisson models are considered which are:

1. $E[y] = e^{\beta_0+\beta_1 x+\beta_2 t}$
2. $E[y] = \beta_0 + \beta_1 x + \beta_2 t$
3. $E[y] = (\beta_0 + \beta_1 x + \beta_2 t)^2$

For each model:

1. (a) Fit a GLM and see if there is an statistically significant relation between the number of visits and the explanatory variables.
   (b) Does the model proposed fit the data appropriately?
   (c) Does it exists overdispersion/underdispersion? If yes, why?

Choose the model that better fit your data:

1. (a) which model have you chosen? Why?
   (b) Compute the expected number of visits in the case where the pollen level is equal to 1 and 4, combined with temperatures of 10 and 20 degrees.
   (c) Analyse if it is possible to use a model with just one covariate to explain these data. In the case it is possible, which covariate would you choose?

QUESTIONNAIRE

1. *Ap.1 Modelo 1.* The value of β_0 is (3 dec.):
2. *Ap.1 Modelo 1.* The value of β_1 is (3 dec.):
3. *Ap.1 Modelo 1.* The p_{value} of β_2 is (3 dec.):
4. *Ap.1 Modelo 1.* The deviance is (3 dec.):
5. *Ap.1 Modelo 1.* The explanatory variables x and/or t do they have any significant effect on the response variable y?
   (a) Yes, sure they have.
   (b) We are not sure, but we accept that they have.
   (c) No, sure they don't.
   (d) We are not sure, but we say that they don't.
6. *Ap.1 Modelo 1.* The log-likelihood value is (3 dec.):
7. *Ap.1 Modelo 1.* The Φ parameter estimation using the Generalized Pearson statistic is (3 dec.):
8. *Ap.1 Modelo 1.* Based on the Φ estimation. Does it exist overdispersion or underdispersion?
   (a) Yes, it exist.
   (b) We are not sure, but we accept that it exists.
   (c) No, sure it doesn't exists.
   (d) We are not sure, but we say that it doesn't exists.
9. *Ap.1 Modelo 1.* From the plots performed to check the goodness-of-fit of the model. Is it possible to say that the model fits appropriately the data?
   (a) Yes
   (b) No
10. *Ap.1 Modelo 2.* The value of β_0 is (3 dec.):
11. *Ap.1 Modelo 2.* The value of β_1 is (3 dec.):
12. *Ap.1 Modelo 2.* The p_{value} of β_2 is (3 dec.):
13. *Ap.1 Modelo 2.* The deviance is (3 dec.):
14. *Ap.1 Modelo 2.* ? The explanatory variables x and/or t do they have any significant effect on the response variable y?
    (a) Yes, sure they have.

(b) We are not sure, but we accept that they have.

(c) No, sure they don't.

(d) We are not sure, but we say that they don't.

15. *Ap.1 Modelo 2.* The value of the log-likelihood is (3 dec.):

16. *Ap.1 Modelo 2.* The Φ parameter estimation obtained by means of the Generalized Pearson statistic is (3 dec.):

17. *Ap.1 Modelo 2.* Based on the Φ estimation. Does it exist overdispersion or underdispersion?

    (a) Yes, sure it exists.

    (b) We are not sure, but we accept that it exists.

    (c) No, sure it doesn't exists.

    (d) We are not sure, but we say that is doesn0t exists.

18. *Ap.1 Modelo 2.* From the plots performed to check the goodness-of-fit of the model. Is it possible to say that the model fits appropriately the data?

    (a) yes

    (b) No

19. *Ap.1 Modelo 3.* The value of β_0 is (3 dec.):

20. *Ap.1 Modelo 3.* The value of β_1 is (3 dec.):

21. *Ap.1 Modelo 3.* The p_{value} of β_2 is (3 dec.):

22. *Ap.1 Modelo 3.* The deviance is (3 dec.):

23. *Ap.1 Modelo 3.* The explanatory variables x and/or t do they have any significant effect on the response variable y?

    (a) Yes, sure they have.

    (b) We are not sure, but we accept that they have.

    (c) No, sure they don't.

    (d) We are not sure, but we say that they don't.

24. *Ap.1 Modelo 3.* The log-likelihood value is (3 dec.):

25. *Ap.1 Modelo 3.* The Φ parameter estimation obtained by means of the Generalized Pearson statistic is (3 dec.):

26. *Ap.1 Modelo 3.* Based on the Φ estimation. Does it exist overdispersion or underdispersion?

(a) Yes, sure it exists.

(b) We are not sure, but we accept that they have.

(c) No, sure they don't.

(d) We are not sure, but we say that they don't.

27. *Ap.1 Modelo 3.* From the plots performed to check the goodness-of-fit of the model. Is it possible to say that the model fits appropriately the data?

    (a) Yes

    (b) No

28. *Ap.2 a).* Which model have you chosen?

    (a) Model 1

    (b) Model 2

    (c) Model 3

29. *Ap.2 a).* which model has an smaller log-likelihood value?

    (a) Model 1

    (b) Model 2

    (c) Model 3

30. *Ap.2 a).* In which model do the diagnosis plots look better?

    (a) Model 1

    (b) Model 2

    (c) Model 3

31. *Ap.2 b).* The estimated number of visits if x =1 and $t = 10$ is (1 dec.):

32. *Ap.2 b).* The estimated number of visits if x =4 and $t = 10$ is (1 dec.):

33. *Ap.2 b).* The expected number of visits if x =4 and $t = 20$ is (1 dec.):

34. *Ap.2 c).* Which is the value of the VIF associated to t? (1 dec.)

35. *Ap.2 c).* Which is the p_{value} of the test that compares the chosen model with the one obtained once the explanatory variable t has been eliminated from the same model?