# PIE2. Second Deliverable. 2018-19

## Note:

In this second deliverable you have to perform the analysis that is explained in the STATEMENT section.
You do not have to deliver any report or file, you have just to answer the questions that appear in the QUESTIONNAIRE section by means of ATENEA.
You have as many opportunities as you want, in the sense that if you change your opinion and you want to change the answer, you can do it.

## STATEMENT

As in the first deliverable, we have the hight, $\boldsymbol{H}$, of ficus plants as well as the number of days they have been planted, $\boldsymbol{Days}$.

As we have seen in the first delivery, it constitutes an example of data such that requires to transform the response variable by means of the logarithm, and to consider the regression $log\,(H) \sim \alpha + \beta \cdot Days$. But the model doesn't verify the homoscedasticity hypothesis.

In all the exercise the significance level is set to be equal to $\alpha = 0.05$.

The models that we are going to work with are defined in what follows. All of them are generalized linear models with linear predictor $\alpha + \beta \cdot Days$:

$\boldsymbol{ModA}$: The model of Normal distribution with link function= $\sqrt{\phantom{x}}$.

$\boldsymbol{ModB}$: Response with a Gamma distribution and link function= log.

$\boldsymbol{ModC}$: Response with variance function=mu and link= log.

For each one of the models presented, we are going to focus in the goodness-of-fit of the model and in checking if the model hypothesis are verified, in particular special attention will be made in the variance function. Moreover, we are going to be interested in predicting the hight of the ficus when planted ($Days = 0$) and after 105 and 150 days of being planted.

For each one of the models:

1. Fit the data, compute the parameter estimations and interpret them.

2. Perform the residual analysis, using the Pearson residuals. In particular plot the Pearson residuals vs fitted values (type linear predictor). Also with the Pearson residuals, perform the standardized residual analysis.

3. Let us define $FDays$ as the variable $Days$ considered as a Factor. This can be done, because there are sufficient number of observations for each value of the $Days$ variable. For all the models, compare the fits obtained with the ones obtained if at each model it is added the variable $FDays$ as an extra explanatory variable. We are asking you to compare the two models by means the F anova test `anova(first model,second model,test=''F'')`.

4. To see if the variance function of the corresponding model may be assumed, to each of the three models:

   (a) Specify the variance function.

   (b) Perform the plot $\sqrt{|Pearson\, residual|}\, versus\, fitted\, values.$

   (c) Perform the Levene's test to compare the variances of the Pearson residuals in the different groups.

5. Estimate the mean and the variance of to the variable $H$ when $Days$ is equal to $0, 105$ and 150. That is, estimate:

$$E(H|Days = a) \text{ and } Var(H|Days = a),$$

   for $a = 0$, 105 and 150.

6. Compare the results obtained with the three models considered.

# QUESTIONNAIRE

1. *Model: ModA Ap: 1.* The estimation of the dispersion parameter obtained with the Pearson statistic is (3 dec.):

2. *Model: ModA Ap: 2.* Which is the first Day such that the standardized Pearson residual is larger than 2? answer zero in case where none of the residuals has a value larger than 2.

3. *Model: ModA Ap: 3.* The *p-value* of the test is equal to (4 dec.):

4. *Model: ModA Ap: 4.* The value of the contrast statistic associated to the Levene's test is equal to (3 dec.):

5. *Model: ModA Ap: 5.* The estimation of $S\left[H_{|Days=105}\right]$ is (3 dec.):

6. *Model: ModB Ap: 1.* The value of $\hat{\beta}$ is (3 dec.):

7. *Model: ModB Ap: 1.* The estimation of the scale parameter obtained with the Pearson statistic is (3 dec.):

8. *Model: ModB Ap: 2.* Which is the last day in which the |Pearson standardized residual| is larger than 3? Answer 0 if it never gets larger than 3.

9. *Model: ModB Ap: 3.* Does the model appropriately fit the mean of $H$?

   (a) "Yes", sure.
   (b) We are not very sure about it, but we accept "yes".
   (c) "No", sure.
   (d) We are not very sure about it, but we accept "no".

10. *Model: ModB Ap: 4.* Does it exist homoscedasticity on the Pearson residuals obtained?

    (a) Yes, it exists
    (b) We are not very sure about it, but we do not reject that it exists.
    (c) No, sure it doesn't exist.
    (d) We are not very sure about it, but we do not reject that it doesn't exist.

11. *Model: ModB Ap: 5.* The estimation of $S\left[H_{|Days=150}\right]$ is (3 dec.):

12. *Model: ModC Ap: 1.* The residual deviance is (2 dec.):

13. *Model: ModC Ap: 2.* Which is the last day in which the |Pearson standardized residual| is larger than 2? Answer 0 if it never gets larger than 2.

14. *Model: ModC Ap: 3.* Does the model appropriately fit the mean of $H$?

    (a) "Yes", sure.

(b) We are not very sure about it, but we accept “yes”.

(c) “No”, sure.

(d) We are not very sure about it, but we accept “no”.

15. *Model: ModC Ap: 4.* Does it exist homoscedasticity on the Pearson residuals obtained?

    (a) Yes, it exists

    (b) We are not very sure about it, but we do not reject that it exists.

    (c) No, sure it doesn’t exist.

    (d) We are not very sure about it, but we do not reject that it doesn’t exist.

16. *Model: ModC Ap: 5.* The estimation of $Var\left[H_{|Days=105}\right]$ is (3 dec.):

17. Ap: 6. Based on the ratio $\frac{Residual\,deviance}{Null\,deviance}$, which of the three models do you thing it is the best?

    ModA ModB ModC

18. Ap: 6. Based on de Levene Test (*Ap: 4*), which of the three models do you think that better verifies the variance function?

    ModA ModB ModC

19. Ap: 6. Which of the three models fits better $Var\left(H|_{Days=0}\right)$?

    ModA ModB ModC

20. Ap: 6. Which of the three models fits better $Var\left(H|_{Days=150}\right)$?

    ModA ModB ModC