Diagnosis

Francesc Carmona

29 de marzo de 2019

Los ejercicios con (∗) son opcionales, con (∗∗) además son difíciles.

Ejercicios del libro de Faraway

1. (Ejercicio 1 cap. 6 pág. 97)

   Using the `sat` dataset, fit a model with the total SAT score as the response and expend, salary, ratio and takers as predictors. Perform regression diagnostics on this model to answer the following questions. Display any plots that are relevant. Do not provide any plots about which you have nothing to say. Suggest possible improvements or corrections to the model where appropriate.

   (a) Check the constant variance assumption for the errors.

   (b) Check the normality assumption.

   (c) Check for large leverage points.

   (d) Check for outliers.

   (e) Check for influential points.

   (f) Check the structure of the relationship between the predictors and the response.

2. (Ejercicio 2 cap. 6 pág. 97)

   Using the `teengamb` dataset, fit a model with `gamble` as the response and the other variables as predictors. Answer the questions posed in the previous question.

3. (Ejercicio 3 cap. 6 pág. 97)

   For the `prostate` data, fit a model with `lpsa` as the response and the other variables as predictors. Answer the questions posed in the first question.

4. (Ejercicio 4 cap. 6 pág. 97)

   For the `swiss` data, fit a model with `Fertility` as the response and the other variables as predictors. Answer the questions posed in the first question.

5. (Ejercicio 5 cap. 6 pág. 97)

   Using the `cheddar` data, fit a model with `taste` as the response and the other three variables as predictors. Answer the questions posed in the first question.

6. (∗) (Ejercicio 6 cap. 6 pág. 98)

   Using the `happy` data, fit a model with `happy` as the response and the other four variables as predictors. Answer the questions posed in the first question.

7. (∗) (Ejercicio 7 cap. 6 pág. 98)

   Using the `tvdoctor` data, fit a model with `life` as the response and the other two variables as predictors. Answer the questions posed in the first question.

8. (∗) (Ejercicio 8 cap. 6 pág. 98)

   For the `divusa` data, fit a model with `divorce` as the response and the other variables, except year as predictors. Check for serial correlation.

9. (Ejercicio 3 cap. 7 pág. 110)

   Using the `divusa` data:

   (a) Fit a regression model with `divorce` as the response and `unemployed`, `femlab`, `marriage`, `birth` and `military` as predictors. Compute the condition numbers and interpret their meanings.

   (b) For the same model, compute the VIFs. Is there evidence that collinearity causes some predictors not to be significant? Explain.

   (c) Does the removal of insignificant predictors from the model reduce the collinearity? Investigate.

10. (Ejercicio 4 cap. 7 pág. 110)

    For the `longley` data, fit a model with `Employed` as the response and the other variables as predictors.

    (a) Compute and comment on the condition numbers.

    (b) Compute and comment on the correlations between the predictors.

    (c) Compute the variance inflation factors.

11. (Ejercicio 5 cap. 7 pág. 110)

    For the `prostate` data, fit a model with `lpsa` as the response and the other variables as predictors.

    (a) Compute and comment on the condition numbers.

    (b) Compute and comment on the correlations between the predictors.

    (c) Compute the variance inflation factors.

12. (∗) (Ejercicio 8 cap. 7 pág. 111)

    Use the `fat` data, fitting the model described in Section 4.2.

    (a) Compute the condition numbers and variance inflation factors. Comment on the degree of collinearity observed in the data.

    (b) Cases 39 and 42 are unusual. Refit the model without these two cases and recompute the collinearity diagnostics. Comment on the differences observed from the full data fit.

    (c) Fit a model with `brozek` as the response and just `age`, `weight` and `height` as predictors. Compute the collinearity diagnostics and compare to the full data fit.

    (d) Compute a 95% prediction interval for `brozek` for the median values of `age`, `weight` and `height`.

    (e) Compute a 95% prediction interval for `brozek` for `age=40`, `weight=200` and `height=73`. How does the interval compare to the previous prediction?

    (f) Compute a 95% prediction interval for `brozek` for `age=40`, `weight=130` and `height=73`. Are the values of predictors unusual? Comment on how the interval compares to the previous two answers.

Ejercicios del libro de Carmona

1. (∗) (Ejercicio 9.1 del Capítulo 9 página 172)

   Realizar el análisis completo de los residuos del modelo de regresión parabólico propuesto en la sección 1.2 con los datos de tráfico.

2. (∗) (Ejercicio 9.2 del Capítulo 9 página 172)

   Realizar el análisis completo de los residuos de los modelos de regresión simple y parabólico propuestos en la sección 1.2 con los datos de tráfico, pero tomando como variable respuesta la velocidad (sin raíz cuadrada). Este análisis debe justificar la utilización de la raíz cuadrada de la velocidad como variable dependiente.