Technical University of Munich
Department of Mathematics

MA4401 Applied Regression
Homework problem 3

Prof. Donna Ankerst, Stephan Haug

November 28, 2017

Problem H.3

Data on last year's sales (Y), in 100,000s of dollars, in 15 sales districts are given in the file sales.csv. This file also contains promotional expenditures (X1), in thousands of dollars, the number of active accounts (X2), the number of competing brands (X3), and the district potential (X4), coded, for each of the districts.

a) Produce a pairs plot of the data using GGally::ggpairs(). Describe what you see concerning the relation between the four predictors and the response Y.

b) Fit a model, containing all predictor variables, to the data. Do a graphical residual analysis by using autoplot() (depends on the ggfortify package). Are there strong violations of the model assumptions?

c) Test the following hypotheses:

   (i) $\beta_4 = 0$

   (ii) $\beta_3 = \beta_4 = 0$

   (iii) $\beta_2 = \beta_3$

   (iv) $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

   at the 5% level. *Hint:* See Lecture 2b, p. 28ff

d) Write a function test_linht(), which computes the test statistic and p-value for test as given in part c). The input of the function should be the full and reduced model, the degrees of freedom for the additional sum of squares and the degrees of freedom for the residual sum of squares of the full model. The output of the function should be like this

```
test_linht(model_red, model_full, df_add = 1, df_full = 10)

## $test_statistic
## [1] 0.4074726
##
## $p_value
## [1] 0.5375986
```

   Apply your function to the test problem from part c) (ii).

   Hint: Use a list as output of your function. The test statistic and the p-value can then be different elements of the list.

e) Consider the reduced model with $\beta_4 = 0$. Estimate the regression coefficients in this reduced model. Give an interpretation of the estimated coefficients.

f) Using the model in e), obtain a prediction for the sales in a district where X1=3, X2=45, and X3=10. Obtain the corresponding 95% prediction interval.