Technichal University of Munich
Department of Mathematics

MA4401 Applied Regression, Homework problem 5

Prof. Donna Ankerst, Stephan Haug (January 09, 2018)

# Problem H.5

In this problem we will examine data from 27 coral reef heads, Porites lobata (), located in the Great Barrier Reef. Risk and Sammarco (1991) () found that the density of the coral skeletons increases with distance from the Australian shore, due to differences in inshore and offshore environments. The data is contained in the file `coral_reefs.csv` , which can be found on moodle.

**a)** Produce a scatterplot of the data using `ggplot2` functions. Choose different colours for the different reefs. *(1 point)*

**b)** Fit polynomial models of increasing degree to the data. In each step test for a lack of fit of the current model. Stop if you can't reject the null hypothesis anymore. Store all fitted models in a list object. *(3 points)*

**c)** Perform a residual analysis for the last model (the one, which didn't show a lack of fit) from part b). Are there any outliers due to the Cook's D statistic?
Use the function `powerTransform()` from the `car` package to estimate the parameter $\lambda$ of the Box-Cox transformation for this model. Find an appropriate R function to perform a Likelihood-Ratio test of the testproblem

$$H_0 : \lambda = 1 \qquad \text{vs.} \qquad H_1 : \lambda \neq 1 \, .$$

(*Hint:* `help(car::powerTransform)` will be helpful). Based on the test result, decide if a Box-Cox transformation should be applied. *(4 points)*

*Remark:* Use the function `autoplot()` for the residual analysis.

**d)** Compute the adjusted $R^2$ and the AIC for all models stored in the list from part b). For each criteria use the function `sapply()` to compute it for all models simultaneously. Do both criteria also favour the model analysed in part c)? *(2 points)*

*Hint:* Remember that the adjusted $R^2$ is contained in the summary of a linear model.