Data exploration, regression, GLM and GAM course

Highland Statistics Ltd www.highstat.comn

Exercise 4: Bivariate linear regression Bailey data

Data description

See exercise 2. In the data exploration we decided to remove 2 sites. You need to stick to this decision.

Underlying question and task

The aim of this exercise is to get familiar with bivariate linear regression. You would not normally start with bivariate linear regression if there are multiple explanatory variables.

To repeat: This is an R coding exercise and it allows us to get familiar with the output.

The underlying question is whether there is a relationship between density and mean depth. Apply bivariate linear regression to model fish density as a function of mean depth.

Carry out the following steps (you may want to consult the R survival guide for example code).

1. Use the lm function to fit a bivariate linear regression.
2. Explain all the numerical output
    a. What are the estimated parameters?
    b. Are the parameters significantly different from 0 at the 5% level?
    c. How much variation do you explain?
    d. What is the fitted model?
    e. Take a pen and paper and sketch the fitted values.
    f. Now draw the fitted values with the predict function. Add the observed values.
    g. How much variation do you explain?
3. Apply a model validation.
    a. Check homogeneity, normality, influential observations.
    b. Are there any residual patterns?
        i. Plot residuals versus each covariate in the model.
        ii. Plot residuals versus each covariate not in the model.